NCE LIBRARY

AUG 17 1928

Vol. XI, No. 4 *PSYCHOLOGICAL REVIEW PUBLICATIONS* AUGUST, 1928

Journal of Experimental Psychology

EDITED BY

MADISON BENTLEY, UNIVERSITY OF ILLINOIS

HOWARD C. WARREN, PRINCETON UNIVERSITY (*Review*)
RAYMOND DODGE, YALE UNIVERSITY (*Monographs*)
S. W. FERNBERGER, UNIVERSITY OF PENNSYLVANIA (*Bulletin*)
WALTER S. HUNTER, CLARK UNIVERSITY (*Index*)
HERBERT S. LANGFELD, PRINCETON UNIVERSITY, Business Editor

CONTENTS

Behavior with Respect to Short Temporal Stimulus-Forms, II: HERBERT WOODROW 259

The Concept of the Threshold and Heymans' Law of Inhibition, III: LLEWELLYN T. SPENCER AND LOUIS H. COHEN 281

The Phi Gamma Hypothesis: L. L. THURSTONE 293

The Rôle of Articulation in Memorizing: M. C. BARLOW 306

The Prediction from Photographs of Success and of Vocational Aptitude: CARNEY LANDIS AND L. W. PHELPS 313

An Electro-kymograph Used in Recording Vaso-motor Changes: STANLEY B. LINDLEY 325

PUBLISHED BI-MONTHLY
FOR THE AMERICAN PSYCHOLOGICAL ASSOCIATION
BY THE PSYCHOLOGICAL REVIEW COMPANY
PRINCE AND LEMON STS., LANCASTER, PA.
AND PRINCETON, N. J.

Entered as second-class matter, July 1, 1920, at the post-office at Lancaster, Pa.

Psychological Review Publications

of the American Psychological Association

EDITED BY

HOWARD C. WARREN, PRINCETON UNIVERSITY (*Review*)
MADISON BENTLEY, UNIVERSITY OF ILLINOIS (*J. of Exp. Psych.*)
RAYMOND DODGE, YALE UNIVERSITY (*Monographs*)
S. W. FERNBERGER, UNIVERSITY OF PENNSYLVANIA (*Bulletin*)
WALTER S. HUNTER, CLARK UNIVERSITY (*Index*)
HERBERT S. LANGFELD, PRINCETON UNIVERSITY, Business Editor

WITH THE CO-OPERATION OF
MANY DISTINGUISHED PSYCHOLOGISTS

PSYCHOLOGICAL REVIEW

containing original contributions only, appears bimonthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 480 pages.

PSYCHOLOGICAL BULLETIN

containing critical reviews of books and articles, psychological news and notes, university notices, and announcements, appears monthly, the annual volume comprising about 720 pages. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

containing original contributions of an experimental character, appears bimonthly, February, April, June, August, October, and December, the six numbers comprising a volume of about 500 pages.

PSYCHOLOGICAL INDEX

is a compendious bibliography of books, monographs, and articles upon psychological and cognate topics that have appeared during the year. The INDEX is issued annually in May, and may be subscribed for in connection with the periodicals above, or purchased separately.

PSYCHOLOGICAL MONOGRAPHS

consist of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 500 pages.

ANNUAL SUBSCRIPTION RATES

Review: $5.50 (Foreign, $5.75)
Journal: $6.00 (Foreign, $6.25)
Bulletin: $6.00 (Foreign, $6.25)
Monographs: $6.00 per volume (Foreign, $6.30)
Psychological Index: $4.00

Current Numbers: **Review** or **Journal**, $1.00; **Bulletin**, $.60

Review and **Bulletin**: $10.00 (Foreign, $10.50)
Journal and **Bulletin**: $11.00 (Foreign, $11.50)
Review and **Journal**: $10.00 (Foreign, $10.50)
Review, Bulletin, and **Journal**, $15.00 (Foreign, $15.75)
Review, Bulletin, Journal, and **Index**: $18.00 (Foreign $18.75)

Subscriptions, orders, and business communications may be sent direct to the

PSYCHOLOGICAL REVIEW COMPANY
PRINCETON, N. J. OFFICE

Journal of
Experimental Psychology

VOL. XI., No. 4 August, 1928

BEHAVIOR WITH RESPECT TO SHORT TEMPORAL STIMULUS-FORMS

BY HERBERT WOODROW
University of Oklahoma

PART II

EXPERIMENTS BY THE METHOD OF REPRODUCTION

It has been shown, through the investigations summarized in Part I of this paper,[1] that the apparent duration of a short 'empty' interval depends upon the durations of the sounds bounding that interval. These investigations, conducted by the method of Constant Stimuli, demonstrated, as a particularly interesting illusion, that an interval occurring in the temporal form, *l–s*,[2] was considerably overestimated in comparison with an interval occurring in the temporal form *s–l*. The present part of this paper describes experiments by the method of Reproduction, confirming the conclusions reached in Part I with respect to the *l–s* vs. *s–l* illusion. In addition, it presents the results of a survey of reactions to a number of temporal forms, the majority of which are composed of two intervals and three sounds. While the data accumulated with these latter forms are in no single case as extensive as those concerning the forms *l–s* and *s–l*, the results obtained with the various forms harmonize fairly well with each other and throw considerable light on the extent to which the same causes of illusion at work in the isolated *s–l* and *l–s* forms

[1] This JOURNAL, 1928, 11, 167–193.

[2] The way an interval is bounded is described in this paper by designating the 'temporal form' in which the interval occurred. By the form *l–s*, is meant an 'empty' interval, bounded by a long initial sound *l*, and a short terminal sound, *s*.

operate when the forms are made more complex. Such forms as *s–s–l*, *s–l–s* and *s–s–s–l–s*, which are among those studied, constitute an interesting stage intermediate between the simplest temporal combinations (as in Part I) and continuous rhythmical repetition of forms.

Apparatus and Procedure. In the work with the method of Reproduction the subject was asked to listen closely as a single temporal form was given, and then to reproduce the entire form (the sounds and the one or more intervals included between them). The apparatus by which the sounds were produced was the same as that used in the experiments by the method of Constant Stimuli. In reproducing any form the subject strove to make both the sounds and the intervals between them the same as in the stimulus-form. The subjects were instructed to be especially careful, however, to reproduce the intervals as exactly as possible. Sounds were reproduced by holding a key down, and intervals by allowing it to remain up. The key moved through an arc of only 1/16th of an inch, and was so nearly noiseless that the only sensory stimulation produced by its manipulation was that due to the contraction of muscles and the pressure of the finger-tips against the key. It was understood by the subjects, who were with one exception advanced students in psychology, that the time representing a sound would be measured from the instant that contact was made, at the end of the downward movement of the key, to the instant that contact was broken, by release of the pressure of the fingers (whereupon the key was pushed upwards by a spring).

The tapping-key was connected with an electro-magnetic time-marker of the same construction as that used in recording the stimulus-form. The two markers were actuated by currents of the same amperage and wrote alongside the record of a 100 dv. tuning-fork. The stimulating form in every trial, as well as the subject's response, was recorded and measured in the manner customarily employed in the graphic technique. The laboriousness of this technique, involving the counting, in the present instance, of the tuning-fork vibrations of considerable sections of over 500 metres of records, is one reason for the rather scanty and exploratory nature of a considerable portion of this section of the investigation.

No stimulus-form was repeated without the lapse of an interval of from 10 to 15 sec, so that the subject was uninfluenced by whatever rhythmical effects might result from the regularly repeated recurrence of the same form. A verbal signal 'ready' was

called out 1.5 sec before the sounding of the stimulus-form by the experimenter, who was separated from the subject by a closed door. Each presentation of the stimulus-form was followed by a single reproduction. The subject was allowed to begin the reproduction as soon after the stimulus as seemed to him natural.

Critical Estimate of the Method of Mean Error. To what extent the method of Reproduction is adapted to the study of illusions in the estimation of intervals is a question which cannot at present be given a definite answer. Certain it is that one must be extremely cautious in generalizing from data so gathered. It is not self-evident that a subject who, in reproducing an interval, makes it longer than it objectively was has overestimated the objective interval; for it may be that the subject intended to, and believed that he had, set up in reproduction a much shorter interval than he actually had. The same illusions that occur in connection with listening to the temporal forms may occur in connection with their reproduction. There is no apparent reason why an interval between two taps made by the subject may not be judged longer than it was as readily as an interval between two sounds.

In view of the above considerations it is not surprising that, when the subject was asked, upon being given a stimulus-form containing two intervals, not only to reproduce the form but also to state his opinion as to which of the intervals in the form was the longer, the subject's expressed judgment was occasionally at variance with his reproductions. Such contradictions are, it is true, rather exceptional, but were obtained frequently enough to establish the fact that they were not mere accidents. It is possible that such contradiction between judgments of the stimulus-intervals and their reproductions constitute data of considerable interest in connection with certain motor theories of time-perception such as those of Münsterberg[3] and others. These theories imply that our temporal experiences depend upon muscular activities which occur as reactions to temporal stimulus-forms. It is hard to see how temporal judgments can be explained any better when conceived as reactions to muscular reactions made in response to the auditory stimulus-forms than when the temporal

[3] *Beiträge zur experimentellen Psychologie*, 1892, iv, 89 ff.

judgments are conceived as occurring directly in response to the auditory stimulus-forms; and when the muscular reactions, as seen in the taps, disagree with the oral judgments, it would seem a hindrance rather than a help to suppose that the oral judgments are determined by the muscular responses. However this may be, it appears that the subject not infrequently suffers an illusion with respect to his muscular responses. He is sometimes amazed to see the record of his taps, showing how differently he did from what he thought he had done.

All in all, the nature of the relationship between reproductions of temporal forms and oral judgments about them must be regarded as something of a riddle. The characteristics of the reproductions are in themselves, however, of considerable interest. There can be no doubt, moreover, that even with our present limited knowledge, it is frequently possible, from the results of 'reproduction,' to make correct inferences as regards oral judgment. This is proved by the fact that the data to be presented in Table VIII, obtained by the reproduction method, led to the inference that an interval occurring in the temporal form *s–l* must be greatly underestimated as compared with one included in the form *l–s*, an inference which was later definitely confirmed by the experiments conducted by the method already described in Part I.

Reproduction of the Temporal Forms l–s and s–l. Table VIII presents the results obtained when the subjects were asked to reproduce the temporal forms *l–s* and *s–l*. First, five trials were made with one of these forms, the form being sounded once and reproduced once at each trial; then ten trials were made with the other form, and these ten trials were followed by five more with the first form. The table shows, under the heading "Order," whether the temporal form serving as stimulus was *l–s* or *s–l*, and by the letters A, B and C, which have explanatory foot-notes, the duration of the various components of the form. Thus the stimulus-form listed as B under the column headed "Durations," and as *s–l* under that headed "Order," was a form beginning with a short sound (*s*) of .06 sec in length, followed by an interval (*i*) of .50 sec,

TABLE VIII

REPRODUCTION OF THE STIMULUS-FORMS *l–s* AND *s–l*

C.E.$_I$ = constant error in reproductions of the interval

$I_{l-s} - I_{s-l}$ = the difference between the constant errors in the reproductions of the intervals in the forms *l–s* and *s–l*

Subj.	Stimulus-Form: Durations	Order	Reproduction: First Sound	I	Second Sound	C.E.$_I$	$I_{l-s} - I_{s-l}$
Hr.	A^1(*i* = .60)	*l–s*	.721	.665	.316	+.065	
			.037	*.024*	*.008*		+.168
	A "	*s–l*	.235	.497	.983	−.103	*.033*
			.007	*.022*	*.072*		
Mc.	A "	*l–s*	.395	.595	.130	−.005	
			.020	*.006*	*.005*		+.174
	A "	*s–l*	.219	.421	.518	−.179	*.015*
			.006	*.014*	*.021*		
Cg.	A "	*l–s*	.557	.645	.107	+.045	
			.009	*.010*	*.005*		+.126
	A "	*s–l*	.143	.519	.517	−.081	*.016*
			.007	*.012*	*.013*		
Ll.	A "	*l–s*	.635	.485	.200	−.115	
			.044	*.029*	*.013*		−.036
	A "	*s–l*	.192	.521	.559	−.079	*.033*
			.006	*.016*	*.008*		
Hd.	B^2(*i* = .50)	*l–s*	.450	.508	.130	+.008	
			.013	*.029*	*.008*		+.067
	B "	*s–l*	.138	.441	.462	−.059	*.030*
			.006	*.006*	*.025*		
Ms.	B "	*l–s*	.505	.563	.169	+.063	
			.012	*.008*	*.012*		+.231
	B "	*s–l*	.098	.332	.470	−.168	*.011*
			.006	*.007*	*.017*		
Ww.	B "	*l–s*	.485	.637	.068	+.137	
			.021	*.031*	*.007*		+.217
	B "	*s–l*	.141	.420	.570	−.080	*.034*
			.008	*.013*	*.028*		
Kl.	B "	*l–s*	.495	.375	.108	−.125	
			.020	*.009*	*.013*		+.095
	B "	*s–l*	.176	.280	.554	−.220	*.015*
			.003	*.012*	*.009*		
Cg.	C^3(*i* = .80)	*l–s*	.731	.757	.096	−.043	
			.010	*.012*	*.007*		+.130
	C "	*s–l*	.148	.627	.689	−.173	*.016*
			.006	*.011*	*.017*		

Av. = +.130

M. V.$_{av.}$ = .020

[1] A, *l* = .45 sec; *i* = .60 sec; *s* = .07 sec.

[2] B, *l* = .38 sec; *i* = .50 sec; *s* = .06 sec.

[3] C, *l* = .60 sec; *i* = .80 sec; *s* = .10 sec.

and ending with a relatively long sound (*l*) of .38 sec. The averages of the subjects' reproductions of each form are given in the three columns headed "First Sound," "I," and "Second Sound." "First Sound" refers to *l* in case of the form *l–s* and to *s* in the case of *s–l*. The letter, 'I,' stands for the subject's reproductions of the stimulus-interval (*i*). The variability given under each average is the mean variation of the average. In the column headed 'C.E.$_I$' is given the constant error made by each subject in his reproductions of the interval, *i.e.* the value obtained by subtracting the stimulus-interval from the average of a subject's reproductions of the interval (I–*i*). When this difference is positive, it means that the subject made the interval longer than it was, and when negative, the reverse. In the last column is given the difference obtained between the constant errors made by a subject in reproducing the intervals of the two forms *l–s* and *s–l*. When this difference is positive, as it is for all subjects except one, where it is unreliably negative, it means that the subject, in reproducing the forms, made the interval in *l–s* longer than the interval in *s–l*. The variability given under each difference is the mean variation of the difference (between the two average reproduction-values). It should be noted that, while several different stimulus-forms were used, the table pairs the reproductions of each *l–s* form with that of an *s–l* form possessing identical total length and identical values for the durations of *s* and *l*. All values in the table are in terms of seconds.

Table VIII brings out the facts as regards the errors made in reproducing the intervals, the study of which was the chief purpose of the experiment. It is true, however, that the errors in reproducing the sounds are equally striking. These are shown separately in Table IX.

The most important conclusions established by the figures in Tables VIII and IX, concerning reproductions of the temporal forms *l–s* and *s–l*, are the following:

(1) An interval in the stimulus-form *l–s*, when reproduced, is made longer than an equal interval occurring in the stimulus form *s–l*. This conclusion holds of all subjects except one,

TABLE IX

CONSTANT ERRORS IN REPRODUCTION OF SOUNDS

Subj.	Form	Longer Sound Stim.	Reprod.	C.E.	Shorter Sound Stim.	Reprod.	C. E.
Hr.	*l–s* A	.45	.721	+.271	.07	.316	+.246
Mc.	" "	.45	.395	−.055	.07	.130	+.060
Cg.	" "	.45	.557	+.107	.07	.107	+.037
Ll.	" "	.45	.635	+.185	.07	.200	+.130
Av.				+.126			+.068
Hr.	*l–s* B	.38	.450	+.070	.06	.130	+.070
Ms.	" "	.38	.505	+.125	.06	.169	+.109
Ww.	" "	.38	.485	+.105	.06	.068	+.008
Kl.	" "	.38	.495	+.115	.06	.108	+.048
Av.				+.104			+.059
Cg.	*l–s* C	.60	.731	+.131	.10	.096	−.004
Hr.	*s–l* A	.45	.983	+.533	.07	.235	+.165
Mc.	" "	.45	.518	+.068	.07	.219	+.149
Cg.	" "	.45	.517	+.067	.07	.143	+.073
Ll.	" "	.45	.559	+.109	.07	.192	+.122
Av.				+.194			+.127
Hr.	*s–l* B	.38	.462	+.082	.06	.138	+.078
Mc.	" "	.38	.470	+.090	.06	.098	+.038
Ww.	" "	.38	.570	+.190	.06	.141	+.081
Kl.	" "	.38	.554	+.174	.06	.176	+.116
Av.				+.134			+.078
Cg.	*s–l* C	.60	.689	+.089	.10	.148	+.048

Av. C. E. for *l* in *l–s* = +.117 sec
for *l* in *s–l* = +.156 sec
for *s* in *l–s* = +.056 sec
for *s* in *s–l* = +.096 sec
for *l*, both forms = +.136 (31 per cent)
for *s*, both forms = +.076 (110 per cent)

who shows a small and unreliable error of the opposite direction. On the average, in all trials of all subjects, the interval in *l–s* is made .13 sec longer than that in *s–l*, a difference which is 6.5 times its mean variation. Since on the average the duration of the reproductions of the interval in *s–l* was .451 sec, it follows that the interval in *l–s* was made 29 per cent longer than that in *s–l*. This result appears to be in fairly close agreement with that obtained by the method of Constant Stimuli, namely, that the interval in *s–l*, when compared with one of .50 sec in the form *l–s*, had to be made 33 per cent [4] longer before being judged equal (longer as well as shorter in one-half the trials).

[4] See Part I, p. 174.

(2) Reproductions of the interval in *s–l* tend to be shorter than the stimulus-interval (true of all subjects) while reproductions of the interval in *l–s* tend to be sometimes longer and sometimes shorter than, and on the average about equal to, the stimulus-interval.

(3) The reproductions of the sounds show a very decided tendency to exceed in length the corresponding sounds of the stimulus-forms. General averages show the reproduction of the long sound to be .136 sec, or 31 per cent, too long; and that of the short sound to be .076 sec, or 110 per cent, too long. In the case of both sounds, this tendency towards too long a reproduction is greater with the form *s–l* than with the form *l–s*. To make sure that the enormous over-reproduction of the short sound was not due to any impossibility of pressing the key down for such a short period, trials were made with all subjects to determine the shortest contact which they could produce. No subject was found who could not hold down the key for a period at least as short as .02 sec, *i.e.*, for a period only one third as long as that required correctly to reproduce the shortest sound used in any of the stimulus-forms.

Some suggestions towards an explanation of the three general results mentioned above may not be out of place. The first two, bearing on the reproduction of the intervals, can be understood on the assumption that the subject 'hears' or 'perceives' the intervals falsely—the interval in *l–s* as either too long or too short and that in *s–l* always as too short—and that he then produces intervals substantially of the same length as he 'perceives' them to have. What one is to understand by the phrase 'perception of an interval' is a question which depends upon one's general psychological point of view and, more particularly, upon one's hypotheses concerning the nature of temporal judgments, disputable matters which will not here be discussed. An explanation of why the perceptions of the intervals, as expressed in the judgments, are illusory, has been attempted in Part I.[5]

Excessive length of the reproductions of the sounds was noted in the case of all the temporal forms employed in the

[5] See pp. 192-193.

present investigation. A possible greater slowness in stopping a contraction (pressing down) than in initiating one,[6] may have something to do with the matter. The overestimation of filled as compared with empty intervals may also be a factor. Filled intervals of the length of the longer sounds here employed are overestimated as compared with empty intervals.[7] Further facts appear to be needed concerning the effect of 'filling' in the case of very short intervals. On the whole, however, probably of greatest value as a clue towards explaining the lengthening of the sounds in reproduction is the observation, first made by Meumann, that in such short temporal forms as those which are under consideration the bordering sensations dominate; that the sounds, rather than the intervals, loom large in the impression made by the total form.[8] It seems probable, then, that discrepancies between stimulus and reproduction, both in the case of the sounds and of the intervals, are for the greater part to be explained in terms of factors affecting the perception of the stimulus-form rather than in terms of factors connected merely with the execution or perception of the reproducing movements.

Reproduction of Complex Stimulus-Forms. The experiments on the reproduction of the forms *l–s* and *s–l*, together with the data reported in Part I on the judgments rendered when the intervals in these forms are compared, afford a fairly complete picture of the responses made to these simple isolated forms. It is an interesting question to what extent the characteristics attaching to these responses can still be traced when these forms are added to, so that they become portions of larger configurations. With this question in mind, experiments were conducted with several forms containing three sounds and with one form containing four sounds. The combination of two short sounds and one long one was studied in the three forms *l–s–s*, *s–l–s* and *s–s–l*. The other combinations investigated were the forms *l–s–l* and *s–s–l–s*. These

[6] For an investigation bearing on this possibility, see R. Sollier, Le temps de reaction à l'arrêt, *J. de psychol.*, 1926, *23*, 980–1002.

[7] See Benussi, *Psychologie der Zeitauffassung*, 1913, 460 ff.

[8] E. Meumann, Beiträge zur Psychologie des Zeitsinnes, *Philos. Stud.*, 1894, *9*, 264–306.

five forms represent an interesting stage between a single rhythmical measure such as *l–s* or *s–l* and a continuous rhythmical series in which *s* and *l* continue indefinitely in alternation.[9]

It should not be assumed, simply because combinations of three or four sounds are intermediate in length between two-sound combinations and continuous rhythmical series, that the factors determining behavior with respect to these intermediate forms are less complex than those determining reactions to continuous rhythmical series. There is some evidence from experiments on comparison of temporal intervals that responses to these intermediate forms are dependent upon a more complex set of factors than are responses to continuous rhythms.[10] The intermediate forms appear to be more ambiguous, that is 'the way they are apprehended'—as inferred from the subjects' verbal reports—is more variable. How the subject will group the sounds is more difficult to predict. In a two-sound combination, like *l–s*, whatever may be the possibilities in the way of variation in the manner of apprehension, only one grouping is possible, namely, *l–s*. In the case of a combination like *s–l–s*, however, there are three possible groupings, namely, $|s–l|–s|$, $|s–|l–s|$, and $|s–l–s|$. In reproductions of forms of this sort it is to be expected that greater individual differences will be found than with the simpler and less equivocal combinations.

The results obtained with combinations of two short sounds and one long one are given in Tables X, XI and XII. Table X presents the results obtained with the form $s_1–s_2–l$. The sounds in this form were so spaced that the first interval—that between s_1 and s_2—was longer than the second interval, that between s_2 and *l*. The reason for this was that otherwise (*i.e.* with equal intervals) the onsets of the three sounds, making a perfectly regular sequence of three beats, might so attract the subject's attention that he would ignore the last

[9] Continuous rhythmical series have also been studied. The results will be reported in a separate paper, in which will be discussed the bearing upon the theory of rhythm of both the data on rhythmical series and those on such short and, in a sense, fragmentary forms as are dealt with in both Parts of the present paper.

[10] See Benussi, *op. cit.*, pp. 99 ff.

sound. It was thought that by making the intervals unequal, and their reproduction therefore presumably more difficult, a better attention to the entire form would be secured. Whether or not this hypothesis be correct, it is a possibility which must be reckoned with that in hearing a series of three sounds the subject may tend to 'regularize them,' *i.e.* to fit them into the schema of a regular sequence of three evenly spaced taps. Indeed, it is not certain that such a tendency may not in part explain the discrepancies between the stimulus-form actually employed, with uneven intervals, and the reproductions of this form, in which the intervals were usually made much more nearly equal than in the stimulus-form.

TABLE X

REPRODUCTION OF THE STIMULUS-FORM $s_1 - s_2 - l$

s_1 and s_2 (each) = .08 sec
l = .50 sec

	Stimulus		Reproduction		$C.E._{I_1}(\%)$	$C.E._{I_2}(\%)$	$C.E._{I_2}-C.E._{I_1}$
Subj.	i_1	i_2	I_1	I_2	$\frac{I_1 - i_1}{i_1}$	$\frac{I_2 - i_2}{i_2}$	
Ww.	.90	.70	.762 *.014*	.796 *.013*	−15.3	+13.7	+29.0%
Cg.	.95	.70	.772 *.026*	.770 *.019*	−18.7	+10.0	+28.7%
"	.95	.45	.644 *.012*	.556 *.011*	−32.3	+23.6	+55.9%
"	.85	.45	.667 *.012*	.485 *.014*	−21.5	+ 7.8	+29.3%
"	.85	.70	.721 *.018*	.768 *.010*	−15.2	+ 9.7	+24.9%
Kl.	.80	.65	.529 *.008*	.529 *.007*	−33.9	−18.6	+15.3%
Ms.	.80	.60	.661 *.009*	.572 *.008*	−17.4	− 4.7	+12.7%
Mc.	.85	.60	.618 *.020*	.391 *.019*	−27.3	−34.8	− 7.5%
Av.	.869	.606	.672	.608	−22.7	+ 0.8	+23.5%

Average length of L = .53; C.E. = +.030; $M.V._{av.}$ = .031.
Average length of S_1 = .12; C.E. = +.040; $M.V._{av.}$ = .011.
Average length of S_2 = .11; C.E. = +.030; $M.V._{av.}$ = .007.

In Table X the intervals of the stimulus-form are listed under the columns headed i_1 and i_2 and the corresponding intervals of the reproduction under I_1 and I_2. The values

given under the latter headings are averages based on ten trials. The mean variation of the average is given under each average. The difference between the stimulus-interval and the corresponding average reproduction is given as a percentage of the stimulus-interval in the columns under C.E.(%). The last column shows the difference in the constant errors for the two intervals, so calculated that when the difference is positive it signifies that the second interval was made longer, relative to its length in the stimulus-form, than was the first interval.

The chief points brought out by Table X may be stated as follows:

(1) Without exception the first interval (s_1–s_2) is made too short by the subject's taps, whereas the second interval (s_2–l) is sometimes made too long and sometimes too short, so that on the average it has practically the same length as in the stimulus-form.

(2) Four of the five subjects shorten the interval s–s, relative to its length in the stimulus-form, more than they do the interval s–l, so that the relative constant error for the latter minus that of the former is, in all instances but one, strongly positive. It is rather questionable whether this result permits one to say that the interval in s–s is underestimated in comparison with the interval in s–l, as was clearly shown to be the case in an experiment reported in Part I (Table VII). It is not obvious that errors made in reproducing the intervals can be taken as measures of errors in the 'estimation' or the judgment of the length of those intervals. It is nevertheless true that the errors made in the reproduction of the intervals are in agreement with the errors of judgment disclosed by the data of Table VII.

(3) On the average, the reproductions of all three of the sounds are longer than the stimulus-sounds. All subjects but one (Mc) showed this characteristic error.

The results obtained on reproduction of the forms l–s–s and s–l–s are shown in Tables XI and XII, respectively. The variabilities given in these tables, are, as usual, the mean variations of averages of ten trials. The durations of the

TABLE XI

REPRODUCTION OF THE STIMULUS-FORM l–s_1–s_2

l = .43 sec; s_1 and s_2 = .06 sec

	Stimulus		Reproduction		C.E.$_{I_1}$(%)	C.E.$_{I_2}$(%)	C.E.$_{I_2}$–C.E.$_{I_1}$
Subj.	i_1	i_2	I_1	I_2	$\frac{I_1 - i_1}{i_1}$	$\frac{I_2 - i_2}{i_2}$	
Bs.	.55	.55	.663	.386	+21.0	−29.8	−50.8%
			.027	*.026*			
Cg.	.90	.90	.866	.744	− 3.8	−17.3	−13.5%
			.058	*.029*			
"	.75	.75	.745	.747	− 0.7	− 0.4	+ 0.3%
			.044	*.084*			
Kl.	.60	.60	.462	.524	−23.0	−12.7	+10.3%
			.009	*.022*			
Ms.	.65	.65	.575	.621	−11.5	− 4.5	+ 7.0%
			.016	*.013*			
"	.60	.95	.636	.752	+ 6.0	−20.8	−26.8%
			.014	*.020*			
Kl.	.60	1.05	.471	.722	−21.5	−31.2	− 9.7%
			.010	*.013*			
Av.	.664	.779	.634	.642	− 4.8	−16.7	−11.9%

Av. length of L = .549; C.E. = +.119; M.V.$_{av.}$ = .037.
Av. length of S_1 = .129; C.E. = +.069; M.V.$_{av.}$ = .021.
Av. length of S_2 = .152; C.E. = +.092; M.V.$_{av.}$ = .028.

TABLE XII

REPRODUCTION OF THE STIMULUS-FORM s_1–l–s_2

Duration of s = .07 sec

	Stimulus			Reproduction		C.E.$_{I_1}$(%)	C.E.$_{I_2}$(%)	C.E.$_{I_2}$–C.E.$_{I_1}$
Subj.	l	i_1	i_2	I_1	I_2	$\frac{I_1 - i_1}{i_1}$	$\frac{I_2 - i_2}{i_2}$	
Ms.	.37	.60	.60	.358	.473	−40.3	−21.2	+19.1%
				.018	*.026*			
Kl.	.37	.60	.60	.394	.564	−34.3	− 6.0	+28.3%
				.010	*.009*			
Cg.	.48	.80	.80	.770	.758	− 3.8	− 5.3	− 1.5%
				.024	*.033*			
"	.37	.40	.40	.444	.487	+11.0	+21.8	+10.8%
				.012	*.015*			
"	.37	.65	.85	.563	.912	−13.4	+ 7.3	+20.7%
				.022	*.036*			
Av.	.39	.61	.65	.506	.639	−16.2	− 0.7	+15.5%

Av. length of S_1 = .012; C.E. = +.052; M.V.$_{av.}$ = .038.
" " " L = .482; C.E. = +.092; M.V.$_{av.}$ = .054.
" " " S_2 = .011; C.E. = +.041; M.V.$_{av.}$ = .024.

intervals in the stimulus-forms are given in the columns headed by small letters (i_1 and i_2) and the average duration of the corresponding reproduction-intervals in those headed by capital letters (I_1 and I_2). Similarly, the stimulus-sounds are represented by *l* and *s* and their reproductions by *L* and *S*.

Tables XI and XII, especially the former, reveal pronounced individual differences. These differences may be due in part to the fact that different lengths of intervals were used with different subjects, and in part to 'accidental variation.' On the whole, however, the data of Tables XI and XII clearly establish the conclusion that the forms used as stimuli determine only to a degree the nature of the subjects' responses.

In addition to this fact of the ambiguity of these stimulus-forms, a second conclusion may be drawn, pertaining to the effect of the stimulus as such, in so far as such effect, in abstraction from the subjects' individual peculiarities, may still be observed, *i.e.* in so far as the reproductions show what might be termed a correlation with the stimulus-form, other factors held constant. The data permit the conclusion that to such degree, slight as it may be, as the errors of reproduction may be regarded as directly dependent upon the stimulus-form, they are in harmony with the already (Part I) ascertained overestimation of the interval *l–s* as compared with either the intervals *s–s* or *s–l*. Thus, in the reproductions of the form *l–s–s* (Table XI), the first interval *l–s* was, on the average, made 11.9 per cent longer than the interval *s–s*; and in the reproductions of the form *s–l–s*, the interval *l–s*, in this case the second interval, was made 15.5 per cent longer than the interval *s–l*. A third conclusion to be drawn from Tables XI and XII is that the reproductions of the forms *l–s–s* and *s–l–s* show the usual tendency for the taps representing the sounds to exceed in length the corresponding sounds of the stimulus-forms.

The remaining two forms which were studied, *l–s–l* and *s–s–l–s*, yielded fairly uniform behavior on the part of the various subjects. The results for *l–s–l* are shown in Table XIII and for *s–s–l–s* in Table XIV. The abbreviations used in these tables need no explanation since they have the same

significance as in the preceding tables. All measures of variability given are the mean variations of averages or of the difference between averages. Each average is the result of ten measurements.

TABLE XIII

REPRODUCTION OF THE STIMULUS-FORM l_1–s–l_2

Length of l_1 and l_2 = .42 sec
Length of s = .06 sec

Subj.	Stimulus i_1 and i_2	Reproduction I_1	Reproduction I_2	C.E.$_2$ $I_2 - i_2$	C.E.$_1$ $I_1 - i_1$	C.E.$_2$–C.E.$_1$ $I_2 - I_1$
Mc.	.85	.776	.452	−.398	−.074	−.324
		.053	*.030*			*.061*
Cg.	.95	.813	.737	−.213	−.137	−.076
		.030	*.020*			*.036*
Ms.	.65	.494	.451	−.199	−.156	−.043
		.015	*.014*			*.021*
Kl.	.55	.472	.388	−.162	−.078	−.084
		.016	*.015*			*.022*
Av.	.75	.639	.507	−.243	−.111	−.132

Av. L_1 = .579; C.E. = +.159; M.V.$_{av.}$ = .072.
Av. S = .138; C.E. = +.078; M.V.$_{av.}$ = .036.
Av. L_2 = .583; C.E. = +.163; M.V.$_{av.}$ = .093.

TABLE XIV

REPRODUCTION OF THE STIMULUS-FORM s_1–s_2–l–s_3

Length of l (constant) = .57 sec
Length of s_1, s_2 and s_3 (constant) = .05 sec

Subj.	Stimulus i_1 i_2 and i_3	Reproduction I_1	Reproduction I_2	Reproduction I_3	Errors I_3–I_1	Errors I_3–I_2
Kl.	.65	.423	.367	.466	+.043	+.099
		.015	*.005*	*.007*	*.017*	*.009*
Ms.	.45	.257	.296	.445	+.188	+.149
		.015	*.006*	*.032*	*.035*	*.033*
Ms.	.70	.555	.588	.794	+.239	+.206
		.015	*.014*	*.011*	*.019*	*.018*
Bs.	.65	.426	.397	.821	+.395	+.424
		.020	*.022*	*.031*	*.037*	*.037*
Cg.	.70	.594	.701	.793	+.199	+.092
		.016	*.017*	*.021*	*.026*	*.027*
Cg.	.50	.302	.351	.288	−.014	−.063
		.014	*.020*	*.023*	*.027*	*.030*
Cg.	1.15	.854	.956	1.272	+.418	+.316
		.042	*.038*	*.047*	*.063*	*.060*
Av.	.686	.487	.522	.697	+.210	+.175
Av. M.V.		*.020*	*.017*	*.025*	*.032*	*.030*

Av. S_1 = .198; C.E. = +.148; M.V.$_{av.}$ = .012.
Av. S_2 = .169; C.E. = +.119; M.V.$_{av.}$ = .006.
Av. L = .580; C.E. = +.010; M.V.$_{av.}$ = .037.
Av. S_3 = .169; C.E. = +.119; M.V.$_{av.}$ = .029.

In the reproductions of the form *l–s–l*, the tendency of the 'tapped sounds' to be longer than the stimulus sounds is unusually pronounced. In view of this over-emphasis of the sounds, it is perhaps not surprising that both intervals in the reproduction are uniformly made shorter than the stimulus-intervals. The important point, however, is that the first interval—that included in the *l–s* portion of the total form—is uniformly made longer than the second interval—that included in the *s–l* portion. It thus seems reasonable to suppose that the illusion of estimation already demonstrated (Table II) as consisting in the overestimation of the interval in *l–s* as compared with that in *s–l* is reflected by all subjects in their reproductions of the stimulus-form *l–s–l*.

The results obtained with the form *s–s–l–s*—the only form used which contains four sounds—are shown in Table XIV. This form contains three intervals, those included in the part-forms *s–s*, *s–l* and *l–s*. In the reproductions, the interval *l–s* (I_3) was longer, and the intervals *s–s* and *s–l* shorter, than the corresponding stimulus-intervals. In consequence, though the three intervals had the same length in the stimulus-form, the reproductions showed the interval *l–s* (I_3) to be longer than either of the others. The only exception was shown by a subject (Cg) who conformed with the general rule in two out of three cases. A comparison of the lengths given the second interval, *s–l*, with those given the first, *s–s*, reveals greater individual differences; although, on the average, the interval in *s–l* is made slightly the longer. In general, then, the response-intervals set up by the subjects, in the case of the stimulus-form *s–s–l–s*, possess relative lengths as follows: *l–s* greater than either *s–l* or *s–s*, and *s–l* slightly greater than *s–s*. These relationships are in perfect agreement with the illusions of estimation described in Part I, consisting in the overestimation of the interval in the form *l–s* as compared with that in *s–l*, and the overestimation of the latter interval as compared with that in *s–s*.

General Conclusions. The results from each of the stimulus-forms used have now been presented and in each case the chief conclusions justified by them have been stated. It remains

to consider the outcome of the entire work with the method of Reproduction. The following appear to be the principal points of interest.

(1) Marked constant errors occur in the reproductions of every one of the stimulus-forms studied.

(2) The degree of uniformity in the constant errors shown by the subjects varies widely with the various stimulus-forms. Individual differences are smaller with forms composed of two sounds (*l–s* or *s–l*) than with forms composed of several sounds, and among the latter individual differences are most pronounced with the form *l–s–s*.

In no case can it be said either, on the one hand, that the constant errors are independent of the particular stimulus-form used or, on the other hand, that they depend entirely thereon. In other words, in all cases, the nature of the constant errors shown by the reproductions of any stimulus-form is influenced by two factors: first, the nature of the stimulus-form; and secondly, a subjective factor, in reality, no doubt, a complex of factors affecting the reproduction through their influence upon such internal processes as are covered by the concepts of 'set,' subjective grouping, manner of apprehension of the stimulus-form, accuracy of perception of finger movements, initiation and execution of movements, etc. The degree to which these internal processes are independent of the stimulus-form is presumably slight in the case of those forms where the individuals are in close agreement but it is certainly great in the case of those where individual differences are marked.

(3) If individual differences are ignored, and only the general or average results obtained with each form considered, the constant errors in the reproductions of the intervals all accord with certain general principles. This is true, however, only providing that the constant errors for one interval are contrasted with those of another interval occurring in the same total form, or, in the case of the one-interval forms, *l–s* and *s–l*, that the constant errors for these two intervals are contrasted with each other.

Altogether, five complex stimulus-forms, that is, forms

containing more than one interval, have been investigated. These five forms contain a total of thirteen intervals. The data obtained afford seven comparisons of the constant errors for one interval with those for a different interval occurring in the same total stimulus-form. These comparisons reveal that an interval of the type *l–s* is, on the average, reproduced as longer than one of the type *s–l* or *s–s* occurring in the same stimulus-form and that one of the type *s–l* is, on the average, made longer than one of the type *s–s* forming part of the same form. Similarly the one-interval stimulus-form *l–s* is reproduced by a longer interval than is the stimulus-form *s–l*. The constant errors observed are thus consistent as regards direction. As regards the magnitude of the constant errors there is considerable inconsistency a large part of which must be attributed to the admittedly rather low reliability attaching to the present quantitative determinations. All that is implied, when it is said that the constant errors are consistent with each other and with those shown by the experiments on judgment, is this consistency in the direction of the errors. From the exact quantitative point of view, the various observed errors show considerable variation.

It will be observed that in the above discussion of the constant errors in the reproductions of intervals, the constant errors for one interval are contrasted with those of other intervals of the same form rather than with those of other forms. Only in one case is there made an exception to this procedure, namely, in the case of the one-interval forms, *s–l* and *l–s*, the constant errors for which have been compared with each other. The reason for the procedure that has been adopted, which is believed to be the only justifiable one, lies in the fact that it is necessary to distinguish two factors: first, errors in the reproduction of the form as a whole; and secondly, errors in the reproduction of each interval, due to the way that particular interval is bounded. The first factor may show itself in tendencies which characterize the reproductions of all intervals in a stimulus-form, irrespective of the way they are bounded. For example, there may exist in most subjects a tendency to make both the intervals in the form *l–s–l* too short. The cause of this tendency might be the relatively great predominance of sound as compared with interval, which exists in this form. Such a cause is quite a different thing from the pattern of bounding which characterizes each interval occurring in this stimulus-form, *i.e.* the *l–s* type of bounding of the first interval, and the *s–l* type of bounding of the second interval. Other total stimulus-forms might produce in the subjects a tendency to make all the intervals relatively long. The causes of these tendencies due to the total form might often be very obscure. Now the effect of these tendencies must be eliminated if one is to determine the effect of the way the individual intervals are bounded. This can be done by means of various hypothetical assumptions. The simplest and most direct way, however, is to compare the reproductions of any

interval only with other intervals in the same stimulus-form, not only on the average but in the case of each subject. Such is the procedure which has been adopted. The only exception to this statement is, as has been indicated, in the case of the two-stimulus-forms containing but one interval, namely, the forms *l–s* and *s–l*. In these cases the terms 'total form' and 'bounding of a particular interval' become synonymous, since each total form consists of only one interval bounded in one particular way. Even so, the comparison of the constant errors for these two total forms would appear to be unjustified were it not for the fact that they were identical in all respects except type of bounding, that is, identical as regards the duration of both the long and short sounds and the duration of the interval.

(4) In view of the correspondence between the constant errors in the reproduction of the two-sound forms and the illusions, described in Part I, revealed by the judgments concerning the intervals of these same forms, it is a reasonable hypothesis that the causes of these constant errors are the same in both cases—reproduction and judgment. No doubt a natural way of conceiving the matter is to assume that the effect of a certain type of bounding is to cause the interval bounded to seem longer or shorter than the same interval differently bounded, and to suppose that the subject both judges and reproduces the interval in accordance with the length it 'seems' to have. But when one asks what is meant by the 'seeming' length of an interval, one encounters the bewildering differences of point of view characteristic of current psychology. There can be little doubt, however, that substitutes for the concept of 'apparent duration' can be found in hypothetical cerebral or cerebellar processes or in 'implicit' muscle patterns. Whatever may be the best explanatory concepts, it is a fact calling for explanation that the reproductions and judgments show identical characteristics dependent upon the duration of the sounds of the stimulus-forms.

(5) The similarity as regards the direction of the constant errors of the reproductions which exists between forms composed of two sounds and those composed of a larger number of sounds indicates a similarity of cause in the two cases. It shows that certain elements in a stimulus-form may exert an influence dependent upon the nature of these elements as such; an influence, that is, which remains constant in spite of change in the type and the degree of complexity of the total-form into which the element enters.

(6) Lastly, it may be noted that, in the case of all the stimulus-forms investigated, the reproductions show a constant error consisting of excessive lengthening of the taps representing the sounds of the forms. This may be related to the fact that it is the duration of the sounds which determines the constant errors in the reproductions and judgments of the intervals. Both facts are in line with the notion, perhaps rather vague, that, in proportion to their physical length, the sounds are of greater importance than the intervals, in determining a subject's behavior in response to a total-form.

General Summary

Part I. A study was made, by the method of Constant Stimuli, of judgments of the relative length of 'empty' intervals (.28–.92 sec in length), when the intervals are limited by sounds of varying lengths. An interval, together with the sounds limiting it, is referred to as a temporal 'form.' Representing a relatively long sound (.38–.50 sec) by *l*, a relatively short sound (.06–.08 sec) by *s*, one of intermediate or medium length (.29 sec) by *m*, and the interval by a dash, the temporal forms used may be represented as follows: (1) *l–s*, (2) *s–l*; (3) *m–m*; and (4) *s–s*. When the interval in one of these forms was compared with the interval in another form, several striking constant errors of judgment, or illusions, were discovered.

All of these illusions may be summarized by the statement that a long initial sound (the sound preceding the interval) causes the following interval to be judged longer, as does also a long terminal sound (the sound ending the interval); but that the lengthening effect upon the interval exerted by a long initial sound is decidedly greater than that of a long terminal sound. A theoretical explanation of these effects was suggested.

An analysis was made of the effect of order of presentation of the standard and variable intervals. This effect was found to be very peculiar and appeared to require for its explanation the assumption that the subjects' judgments were influenced by what is known as *absolute impression*.

Lastly, in the case of the overestimation of the interval in *l–s* as compared with that in *s–l*, it was found that, with increase in the number of trials up to four hundred, the illusion tended to increase.

Part II. Records were obtained of reproductions of various temporal stimulus forms, the reproductions being made by pressures upon a contact-key. The stimulus-forms used, seven in number, were the following: *s–l*; *l–s*; *l–s–s*; *s–l–s*; *s–s–l*; *l–s–l*; and *s–s–l–s*. Comparisons were made between the reproductions of an interval in any stimulus form with those of another interval in the same form, and between those of the form *l–s* and those of the form *s–l*. Such comparisons revealed, in the case of the reproductions of the intervals of every stimulus-form, marked constant errors which were dependent upon the way the intervals were bounded. The constant errors shown by the various subjects were with some forms quite uniform, while with others the individual differences were so marked as greatly to obscure any general tendencies. If, in spite of individual differences, such general tendencies as were traceable were considered, these were found always to be consistent with each other and with the errors of judgment revealed by the experiments described in Part I.

In particular, the following general tendencies were noted:

1. Reproductions of the interval of the stimulus-form *l–s* longer than those of the interval of the stimulus-form *s–l*.

2. With the form *s–s–l*, a lengthening of the second interval (*s–l*) in comparison with the first (*s–s*).

3. With the form *l–s–s* a lengthening of the first interval (*l–s*) in comparison with the second (*s–s*) (general tendency very unreliable because of individual differences).

4. With the form *s–l–s*, a lengthening of the second interval (*l–s*) in comparison with the first (*s–l*).

5. With the form *l–s–l*, a lengthening of the first interval (*l–s*) in comparison with the second (*s–l*).

6. With the form *s–s–l–s*, a lengthening of the third interval (*l–s*) in comparison with either of the others (*s–s* or *s–l*) and a tendency (unreliable) to lengthen the second interval (*s–l*) in comparison with the first (*s–s*).

All of the above tendencies may be summarized by the schema:

Interval in *l–s* greater than that in *s–l*.
Interval in *s–l* greater than that in *s–s*.

The same schema holds for the illusions demonstrated by the method of Constant Stimuli.

Owing to the agreement in the direction of these constant errors detected in the two forms of behavior, judgment and reproduction, it is suggested that these errors have the same explanation in both cases. Further, since these errors noted in part-forms of complex total-forms agree with those noted in the isolated part-forms, it is inferred that a given pattern of bounding sounds exerts an effect upon the interval bounded which remains constant no matter what the type or complexity of the total-form of which this pattern constitutes an element. This last conclusion does not imply that the total pattern as such does not influence behavior. It emphasizes only the fact that a particular sub-pattern exercises an influence which remains constant irrespective of the total pattern into which it enters and of any effect exerted by the latter.

(Manuscript received Nov. 24, 1927)

THE CONCEPT OF THE THRESHOLD AND HEYMANS' LAW OF INHIBITION

BY LLEWELLYN T. SPENCER AND LOUIS H. COHEN

Yale University

III. The Relation of the Threshold to Estimates of Daily Variation in 'Freshness'

In the preceding papers of this series (5, 6) the variation of the simple visual threshold with the variation of a measure of susceptibility to inhibition by another visual stimulus simultaneously presented has been discussed. Evidence was presented from experiments by Heymans showing the concomitance of such variations for qualitative differences with a single subject, from other experiments by Spencer with stimulus of a constant character for individual variation within a group of 50 subjects and most recently by the present writers for daily variations of a single subject. Heymans' original observation was interpreted by him as indicating the primary importance of inhibitory factors in determining the value of the stimulus threshold. It has been stressed in this series of papers that the succeeding correlations reported by us are not to be directly deduced from Heymans' results but do follow from his theory of threshold determinations. So far, then, all experimental results designed to test Heymans' hypothesis have fully supported his suggestion.

Many other factors, however, are traditionally held to influence the threshold. Of these the matter of fatigue has been considered even from the time of the original generalizations of Weber. A long line of experimenters whose results are summarized by Whipple (9) and evaluated and extended by Fernberger (2) has been concerned with the effect of protracted mental and physical activity upon subsequent measurements of the threshold. The senses of pain, vision and hearing, the use of faradic stimulation, two-point cutaneous

discrimination and lifted-weight comparisons have all been exploited with varying results. A number of investigations have indicated decreases in sensitivity as the aftermath of activities reported to be fatiguing. Almost as many have failed to show any effect upon the threshold. For pain, Swift (7) and Vannod (8) report an *increase* in sensitivity after fatiguing work. Swift explains the increase as a result of increased 'irritability of the whole system.' Binet (1) finds, however, a decrease in sensitivity to pain. Coupled with the hyperalgesia which Vannod reports Vannod also found a *decrease* in cutaneous sensitivity as measured by the two-point threshold after physical work. Decreases in sensitivity, where found, are more pronounced after physical work. Fernberger, for example, finds little or no effect upon the interval of uncertainty in lifted-weight comparisons after mental work but regular increases after physical work. In vision and hearing Graziani (3) reports definite increases in sensitivity after the mental and physical effort of lecturing. He also finds after the same activity an increase in the two-point sensitivity for regions on the face but no change for regions on the hand. His results are rendered more convincing by reason of the fact that after a rest-period the sensitivity returns to normal or below.

Historically regarded the interest in these threshold variations has largely followed the hope that if they could be shown to accompany mental work they might be used as a direct or indirect measure of mental fatigue. For our purpose, however, the distinction between mental and physical work is secondary. We are interested in knowing only whether the thresholds obtained after continued work vary from those obtained previous to such work when the subject was presumably less fatigued. Furthermore, it is of considerable importance to discover in which direction such variations occur if they can be demonstrated as regular consequences of prolonged activity.

If such regularity can be shown, the suggestion of Heymans that the phenomenon of the threshold is almost wholly if not entirely due to the inhibitory effect of intra- and extra-

organic stimuli must be reëxamined. It is possible for instance to suppose that fatigue itself may influence the inhibitory effect of such concomitant stimuli. Perhaps for many reasons such an influence might be supposed to be in the direction of a reduction of susceptibility to such effects. That variations in the effectiveness of the effort of attention may play a rôle in the variations of thresholds during fatigue seems to have been a prominent thought in the minds of numerous investigators, of whom Griesbach (4) is perhaps the earliest. This view might lead us to infer that in the relaxation of attention a condition of less integrated response to numbers of stimuli might arise so that the response to the threshold stimulus, being somewhat more independent of concomitant factors, conceivably might be more successful and, as a result, a lower threshold might be obtained.

While such a line of reasoning is obviously speculative, enough has perhaps been suggested to indicate that the relation of fatigue to the threshold is clearly involved in Heymans' hypothesis and further that the hypothesis is not necessarily opposed to the presence of a relation between the two. The disturbing feature of the foregoing argument, however, lies in the fact that the final statement of the preceding paragraph leads directly to the inference that the effect of fatigue would be in the direction of lower thresholds, whereas a considerable number of those who have reported any relation between fatigue and sensitivity find higher thresholds correlated with fatigue.

Unless a correlation of increased fatigue with *increases* of the threshold can be reconciled with Heymans' generalization, it is obvious that the occurrence of an influence due to fatigue and the direction of such influence constitute a critical case for the testing of his hypothesis. Any immediate decision in this crucial case, however, is rendered dubious by our lack of a direct and certain measure of fatigue. Indeed the very conception of fatigue is so hypothetical that any attempt to reduce the theoretical problem to experimental test must be limited to existing means of estimating fatigue effects and can hardly be regarded as decisive.

The mere use of an artificial fatiguing situation, such as mental or physical tasks of uniform character, as a measure of fatigue seems open to objection. While the uniform labor involved may be regarded as a desirable control, the level of 'freshness' before each threshold measurement is by no means rendered quantitatively certain by this procedure, nor can those variations which occur be measured. No reason exists to suppose that activities outside the experimental period fail to have prolonged and cumulative effect upon the freshness of the subject; indeed subjective evidence rather points to a decided influence from this source. Consequently it seems that a preliminary approach to the problem, if quantitative conclusions are to be drawn, were better arranged so as to provide for some estimate of freshness, even if less uniform, representing the final level at the time of threshold determination and expressing the cumulative effect of extra- as well as intra-experimental activity, rather than to accept the conditions immediately preceding the measurement as the sole source of fatigue.

The simplest method of obtaining such an estimate is obviously a subjective rating for freshness. While subjective estimates by members of a group are regarded as unreliable, in the particular case where a single subject is used throughout the experiment and variation is from day to day rather than from individual to individual, such an estimate is much more likely to acquire a comparability from rating to rating so far as the subjective ranking scale is concerned. Another gross but wholly objective method of estimate might be the hours of sleep during the night preceding the threshold measurement, especially when, as in our case, the experimental sitting comes as the first serious task of the day. Obviously neither of these methods affords a perfect index nor do they necessarily act as checks upon each other. The very inaccuracy of the two methods as means of measurement, however, has the negative virtue of assuring that whatever quantitative relationships are found will appear too low rather than too high.

The procedure used in obtaining the visual thresholds has been outlined in the preceding article of this series. It consisted of the method of Minimal Changes combined

with the method of Right and Wrong Cases (Constant Stimuli) at each intensity. The threshold is defined as the lowest value of stimulus intensity for which 80% correct responses are given in any one of three sets of 10 trials. Thresholds were obtained when the intensity of a neighboring stimulus was equal to 0, to 5, and to 10 units on the inverse square scale of a photometer. The subject was dark-adapted for 15 minutes before commencing the threshold determination. The use of a fixation cross screened with a ruby glass filter and set at the right of both the threshold stimulus and the second or inhibitory stimulus resulted in obtaining the thresholds for peripheral stimulation of the eye. A screen limited the field of the left eye to the threshold stimulus and the fixation cross and the field of the right eye to the inhibitory stimulus (when present) and the fixation cross. Otherwise the visual fields were completely dark.

Before the experimentation each day, the subject (Cohen) rated himself for general freshness on a scale of 0–6 and also specifically for visual freshness on a scale of 0–5, being instructed in the latter case to consider particularly subjective impressions of eye-strain, blur and other visual disturbances. He also reported at the same time the number of hours of sleep during the previous night. Experimentation was conducted each day beginning at about 8:45 A.M. Since these reports were made *previous* to any measurements of sensitivity by the threshold procedure, it is obvious that no conscious or unconscious adjustment of ratings to threshold-values was possible. The possibility of adjusting threshold-values to previously given ratings was limited by conditions, enumerated and discussed in the preceding paper, which, it was concluded, made it practically impossible for the subject or the experimenter to influence the values obtained for the various thresholds. Original preconceptions, were they able to influence the relationship, would obviously have been in favor of increased thresholds in correlation with decreased freshness.

The ratings and liminal measurements for 50 days, consecutive except for the omission of Sundays, were correlated with the result shown in Table I. In previous treatment of these same threshold values in the preceding paper it was found that higher correlations were obtained with the measures of susceptibility to inhibition, derived from the three thresholds obtained each day, when the second 25 days were considered separately. This seemed logical by reason of the practice effect from day to day for the single subject, which, it was suggested, should operate in the direction of more reliable measurements as the procedure, requirements and laboratory situation became more familiar and the adjustment of the subject to these conditions more habitual and stereotyped. Similarly, then, it should be expected, other things being equal, that higher correlations should be exhibited in the present instance, if correlation occurs at all, by limiting the cases to the second 25 days. For this reason the r's for the second 25 days are given in Table I along with those based upon the whole 50 observations.

Table I

Correlation Coefficients between Stimulus Threshold and Three Measures of Freshness obtained with a Single Subject

Measure of Freshness	r with $n = 50$	r with $n =$ 2d 25
(*f*) General Freshness r_{tf}	+.533	+.823
(*v*) Visual Freshness r_{tv}	+.565	+.784
(*s*) Hours of Sleep r_{ts}	+.597	+.778

Before any attempt is made to digest the relations exhibited by the Table it may be pertinent to inquire whether the three measures of freshness are independent and complementary or are largely identical. Inasmuch as the judgment of the subject in this respect is obviously undependable, since he made every effort to keep the types of measures independent, the correlation method furnishes the only means to obtain an answer to our inquiry. Table II gives the intercorrelations between the three measures.

Table II

Intercorrelations between the Three Measures of 'Freshness'

	r with $n = 50$	r with $n =$ 2d 25
General with Visual Fr. r_{fv}	.708	.925
General with Hrs. of Sl. r_{fs}	.773	.895
Visual Fr. with Hrs. of Sl. r_{vs}	.793	.903

These correlations show that, especially during the latter half of the experiment, all three types of measures measured approximately the same thing. This not only means that any one could be used as representative of all but it also gives a surprising reliability to the measures taken individually. Furthermore the correlations show that the subjective rating usually so unreliable is, in this case, very closely in agreement with a perfectly objective measurement, viz., the hours of sleep. This close relationship is highly interesting, psychologically, of and for itself. The values of r_{fs} and r_{vs}, .895 and .903, indicate that what the subject recognizes as 'freshness' and as the 'lack of freshness' are respectively almost identical [1] with the occurrence of maximal amount of sleep the night before and with the lack of it.

[1] Although when $r = .90$ considerable disparity is still indicated, the many sources of unreliability, particularly the small number of class intervals, (6 and 7) favor the interpretation of $r = .90$ as indicating an approach to identity.

Returning now to the coefficients of Table I it is clear that there is a considerable relation between values of the threshold and each of the measures of 'freshness.' This is especially true in the second 25 observations. But what is more striking than the mere size of the coefficients is their sign. All measures agree in showing a *positive* relation between the degree of freshness and the height of the threshold. In other words, the fresher the subject the less sensitive his visual system. This is not only contrary to the findings of many of the authors previously referred to in apparently related investigations but is also opposed to the ordinary common-sense conceptions of freshness and sensitivity. Indeed, for the usual assumption that the threshold is a phenomenon determined by an intrinsic resistance of the sensory mechanism these results seem quite unintelligible. We have already seen, however, that for Heymans' threshold hypothesis the relationship here reported is deducible *a priori*.

Now that actual data are at hand to support our speculations, let us restate the argument. When we regard our own experiences we are inclined to concede that, in a 'fresh' state, we present to environmental stimuli a reacting system responsive to any and all influences and characteristically capable of integrating all this multitude of impressions into an organized and focussed whole. As freshness is reduced, however, we experience a deterioration of this organization, an increased reactivity to stimuli as individual, disintegrated units. In the approach to extreme exhaustion, we speak of ourselves as irritable and distractable. By this we mean that we are not able to inhibit the minor irritants and concentrate upon a single desired focus of activity. In saying this we do not mean that we are more susceptible to inhibition; we are actually less susceptible. The ability to control and to integrate our behavior by inhibition eludes us; we approach a condition where we are at the mercy of each and every stimulus. Now if we express these things in the terminology of Heymans' hypothesis, this is just the state of affairs which would produce such correlations as are here reported. Being, when fresh, most completely integrated and susceptible to the

concentrative principle of inhibition, we, therefore, if the threshold is determined by the susceptibility to inhibition, should expect high thresholds. On the other hand, when freshness is low; when we have had but three hours sleep the night before, as in the extreme case in our observations, then we elude the normal, selective action of central inhibition and should expect low thresholds.

It is true, of course, that it is not always certain in just which direction the disintegrating effect of fatigue will act. Although we have assumed so far that the effect of fatigue is chiefly felt in connection with the inhibitory action of concomitant stimuli, it may well be true under certain situations, with certain degrees of fatigue, or in the case of certain sensory or perceptual fields, that the effect of fatigue may be greater in reducing the effectiveness of the reaction to the stimulating situation itself. This, for instance, might well be true in connection with such moderately complex adjustments as are involved in two-point threshold or lifted-weight discriminations. Nevertheless, this ambiguity of direction of influence does not prevent us from admitting the possibility of fatigue effects of such nature as we have been considering for given sensory modalities and experimental situations.

We need not, however, restrict ourselves exclusively to this somewhat unreliable analysis in terms of experience. If these various measures of freshness are related to the threshold only because they are correlative with variations in susceptibility to inhibition these measures should be as closely, or even more closely, related to measures of such susceptibility. Such measures are provided in the data obtained. As described in the previous papers of this series the ratio of increase in the threshold, produced by a second stimulus, to the value of that second stimulus gives a coefficient of inhibition which varies for a single subject from day to day, although remaining relatively constant during the progress of an experimental period. Furthermore, such coefficients of inhibition vary with the thresholds obtained on the corresponding days; not by algebraic implication but rather in accordance with the principle hypothetically suggested by Heymans to explain the fact of the threshold.

If, then, we correlate such measures of inhibition (k) [2] with the measures of freshness we shall have a statistical picture of the degree of implication of the one in the other. Table III presents coefficients of such correlation.

TABLE III

COEFFICIENTS OF CORRELATION BETWEEN MEASURES OF FRESHNESS AND COEFFICIENTS OF SUSCEPTIBILITY TO INHIBITION

	r with $n = 50$	r with $n =$ 2d 25
General Freshness and k, r_{fk}	.456	.841
Visual Freshness and k, r_{vk}	.592	.806
Hours of Sleep and k, r_{sk}	.668	.783

Comparison of these coefficients with those of Table I shows that the measures of freshness are as closely related to the susceptibility to inhibition as they are to the threshold. Indeed, with the high relation between t—the threshold—and k (.968 for $n = 50$ and .992 for $n = 25$), the values of Table I and III imply that but for the association existing between k and the measures of freshness there would be little association between these latter measures and the threshold. This is shown by the values of the correlations of Table I when the factor k is partialled out. Table IV gives these coefficients.

TABLE IV

COEFFICIENTS OF CORRELATION BETWEEN MEASURES OF FRESHNESS AND THE THRESHOLD WITH SUSCEPTIBILITY TO INHIBITION CONSTANT

	r with $n = 50$	r with $n =$ 2d 25
General Freshness and t, $r_{tf \cdot k}$	+.021	−.172
Visual Freshness and t, $r_{tv \cdot k}$	−.040	−.210
Hours of Sleep and t, $r_{ts \cdot k}$	−.266	+.014

As these coefficients show, the association between the threshold and freshness, when susceptibility to inhibition is eliminated or constant, sinks to very low values and frequently becomes inverse, the latter of which facts accords with the expectation of common sense. None of these coefficients is reliable; even the −.226 with 50 cases is less than three times

[2] Correlations with k in this article are uniformly made with the values designated k' in the preceding article of this series and derived from the formula $k' = \frac{t_{A'} - t_0}{A'}$, as explained in the same paper.

the P.E.$_r$. Similarly, the correlations between the threshold and k for $n = 50$ and for $n = 25$ are decreased only .016 or less by the partialling out of any measures of freshness.

Insofar as the statistical methods may be trusted with such a limited sample of cases the correlations of the last two tables support the argument so far adduced. It is not so easy to reconcile the results of our experiment with the numerous investigations whose evidence, insofar as it indicates a relation between fatigue and the threshold, generally points to a relation opposite in direction to that found by us. Aside from the question of whether such investigations of the past and the present study are really comparable, it is only proper to remark that existing evidence is too conflicting to justify the assertion that a direct relation between fatigue and the threshold actually exists. It has already been suggested that it is not impossible to suppose that outcome will vary with the various sense modalities. The effect of fatigue on the visual threshold seems, heretofore, to have been studied only by Graziani, whose results, as described above, coincide with ours as far as direction of the effect is concerned. No investigation, for any sense modality, so far as we have been able to find, subjects the problem to the correlation method of approach with its implied quantitative technique; nor, as we have suggested, is this surprising since the difficulty of finding a quantitative measure of fatigue makes the approach hazardous. Nevertheless, we feel that these results, obtained by us incidental to our major observations, are sufficiently suggestive to encourage at least a repetition of such an investigation in the interests of verification or, better, the devising of superior techniques for independent attacks upon the problem.

Before leaving the subject of these measures of freshness it may be interesting to inquire whether the conditions of the occurrence of the threshold values, so closely comprehended by the factor of susceptibility to inhibition alone, may not be more completely defined by the inclusion of the measures of freshness as well. Expressed in statistical terms our query becomes: Can the high correlation between t and k (.992) be further increased by including the three additional variables

as in $R_{t(kfvs)}$? Since any increases in r_{tk}, whether for the 25 case sample or for the 50 cases, will necessarily be largely made in minor decimal places, the calculation of a five-variable multiple R cannot be made accurately without carrying an unwieldy number of decimals in the lower order partials. For this reason Rs obtained by eliminating independent variables in varying orders will not always, if ever, agree. We can only state the maxima which such Rs reach. For the 50-case sample $R_{t(kfvs)} = .975$ as against $r_{tk} = .9681$; for the 25-case sample $R_{t(kfvs)} = .9928$ as against $r_{tk} = .9920$. Despite the degree of the relation between these measures of freshness and the values of the threshold as exhibited in Table I their independent significance in conditioning the value of the threshold in addition to the susceptibility to inhibition seems practically zero.

The bearing of these present results on Heymans' hypothesis seems to be twofold. (1) Freshness, a type of factor both popularly and by previous experimentation closely linked with the variation of the threshold, seems to have little if any relation thereto apart from its own implication with the susceptibility to inhibition. Heymans' hypothesis therefore remains plausible. (2) For the first time an independent method of approach to the problem of the conditional relation of the threshold to the susceptibility to inhibition is now provided. Whereas all previous support of Heymans' suggestion comes from correlations with factors themselves derived from threshold measurements,—although correctly so, as we have argued,—here in the measures of freshness and especially in the factor of hours of sleep, an entirely independent, non-visual measurement, is found a correlative of the threshold variation which is at least theoretically capable of being harmonized with his theory, but whose relation to the threshold seems inexplicable, if real, by any other existing view as to the nature of the threshold.

Summary

In connection with previously reported investigations concerned with the relation between values of the visual thresh-

old and measures of susceptibility to inhibition by another visual stimulus, the relation to values of the threshold of (1) subjective rankings for general 'freshness' and, more specifically, (2) rankings for 'visual freshness,' as well as (3) the hours of sleep preceding experimentation is reported. The following conclusions are suggested by our limited data from 50 observations on a single subject.

1. Subjective ratings for freshness by a single individual from day to day are closely related to the number of hours of sleep during the night preceding the rating (r = .79 to .90).

2. Such ratings and also the hours of sleep are related to values of the visual threshold, obtained shortly after the ratings, in the order of +.55 to +.80.

3. The positive relationship found is difficult to explain, if true, upon any other basis than the generalization of Heymans that the threshold is almost wholly if not exclusively due to inhibitory effects of intra- and extraorganic stimuli simultaneously present. Freshness, it is suggested, may conceivably, under certain circumstance, be associated with heightened susceptibility to such inhibition. This is supported by correlations of the order of +.60 to +.80.

4. Partial and multiple correlations indicate that the relationship between either freshness or sleep-measures and the visual threshold is negligible except for the factor of susceptibility to inhibition implicated in both.

5. Importance attaches to the correlation described in (2) for it represents the first independent evidence in favor of Heymans' hypothesis apart from measures of susceptibility to inhibition involving visual threshold measurements.

References

1. Binet, A., *L'Année psychol.*, 1905, 11, 1–37.
2. Fernberger, S. W., *J. Exper. Psychol.*, 1916, 1, 508–532.
3. Graziani, A., *Arch. ital. de biol.*, 1907, 250–260.
4. Griesbach, H., *Arch. f. Hygiene*, 1895, 24, 124–212.
5. Spencer, L. T., *J. Exper. Psychol.*, 1928, 11, 88–97.
6. Spencer, L. T., & Cohen, L. H., *J. Exper. Psychol.*, 1928, 11, 194–201.
7. Swift, E., *Amer. J. Psychol.*, 1900, 11, 312–317.
8. Vannod, Th., *Rev. med. de la Suisse Romande*, 1896, 27, 21 ff; also summarized in detail by Vaschide, N., *Rev. de phil.*, 1905, 6, 428–448.
9. Whipple, G. M., *Manual of mental and physical tests*, Baltimore, 1914, 2d ed., I, 252–254, *et passim*.

(Manuscript received Dec. 1, 1927)

THE PHI GAMMA HYPOTHESIS

BY L. L. THURSTONE

University of Chicago

Some time ago I wrote a note calling attention to the fact that the phi-gamma hypothesis is at best an approximation formula and that it is theoretically incorrect. At that time I was not sure whether teachers of psychophysics generally acknowledged this limitation of the phi-gamma hypothesis since I had not seen any explicit statement of its theoretical limitations. As far as I knew the theoretical inaccuracy of the hypothesis might have been considered obvious by psychophysicists generally. Before the note was published I discovered that the theoretical inadequacy of the phi-gamma hypothesis is not generally acknowledged and in consequence I have extended the note into a more complete statement which takes the form of three theorems and several corollaries concerning the psychometric curve and the phi-gamma hypothesis.

The theoretical limitation of the phi-gamma hypothesis consists essentially in the fact that it requires a constant limen for all values of the stimulus. As is well known, the limen for a large stimulus is usually larger than for a small stimulus. Therefore the phi-gamma hypothesis which requires a constant limen is inconsistent with the generally observed fact that the limen is some increasing function of the stimulus. This increasing function is not necessarily that of Weber's law.

It is true that the phi-gamma curve can frequently be used as a good approximate fit for the psychometric curve; but when it is so used it should be thought of not as a psychophysical hypothesis but rather as any other empirical equation which is used to draw a smooth curve through a set of experimental observations.

In this discussion I limit myself to psychometric experiments which are carried out with only two judgments allowed,

heavier or lighter, and from which the intermediate category of judgment is barred. There is of course a difference of opinion as to whether the intermediate category should be used in psychophysics, but that is a different story. *The present argument concerns the phi-gamma hypothesis for psychometric data obtained with only two categories of judgment.*

My propositions concerning the phi-gamma hypothesis are as follows:

Theorem: If the absolute limen increases as the stimulus increases, then the psychometric curve for two categories of judgment is positively skewed.

Corollary: The only condition under which the phi-gamma curve can fit the psychometric curve is that of a constant absolute limen independent of the stimulus magnitude.

Corollary: If the absolute limen increases as the stimulus increases, then the phi-gamma curve, which is symmetrical, cannot fit the psychometric curve, which is skewed.

Corollary: The phi-gamma hypothesis is incompatible with Weber's law. Both cannot apply to the same set of data.

Theorem: If the absolute limen is proportional to the stimulus magnitude (Weber's law), then the positive skewness of the psychometric curve increases with the coarseness of the perceptual process.

Corollary: The necessary error in fitting the phi-gamma curve to psychometric data increases with the relative limen of the perceptual process. The error of the phi-gamma hypothesis is large for coarse discrimination. It becomes negligible for keen discrimination.

Theorem: If Weber's law is satisfied by the psychometric data, *i.e.* if the absolute limen is proportional to the stimulus, then the psychometric curve becomes Gaussian when plotted against the logarithm of the stimulus magnitude.

For the sake of completeness we shall first define the terms used in the theorems and their corollaries. The limen has been described as the absolute amount by which a stimulus must be increased in order that the increase shall be noticeable. With the proverbial lifted weights taken as an example,

the limen for a weight of one hundred grams is the number of grams by which this weight must be increased in order that the increase shall become noticeable. As is well known, the limen is larger for heavy weights than for light.

The above rough description reads as though one experimented with a single weight which is increased in magnitude until the increasing weight becomes noticeably heavier. But that is not the experimental procedure by which the limen is determined. The psychometric function refers to the constant method in which *two* stimuli, *two* weights, are presented and the subject decides which of them is the heavier. Hence the limen is described as the absolute difference between *two* weights which is noticeable.

In order to reduce the term 'noticeable' to something measurable, the noticeability is defined usually as that difference between *two* stimuli which is correctly discriminated in 75% of the attempts. Then *the limen is more accurately defined, with reference to the constant method, as that absolute difference between two stimuli which is correctly discriminated in 75% of the attempts.*

The two terms *standard stimulus* and *variable stimulus* are not infrequently misinterpreted. When the subject is presented with two stimuli, either simultaneously or in succession, his task is to discriminate between them by deciding which is the heavier, larger, or stronger, as the case may be. It should be carefully noted that the subject in his task of discrimination between the two stimuli has absolutely no concern with the distinction between the standard stimulus and the variable stimulus. In other words, this distinction between standard and variable is an arbitrary one introduced later by the experimenter merely for convenience in tabulating and analyzing the records. As far as the subject is concerned, the ideal situation is one in which he is faced with two stimuli and he decides which of them is the stronger.

When every stimulus has been paired with every other stimulus in the series several hundred times, it is possible to tabulate the frequency with which each pair of stimuli is correctly discriminated. The experimenter may then tabu-

late in a single column all the pairs in which any particular stimulus A is a member. These records are then said to have stimulus A as a standard. The same thing can of course be done for each of the other stimuli taken in turn as a standard. It is therefore clear that if the judgment "A is greater than B" is returned in (let us say) 82% of the attempts, then this experimental result belongs equally well in the column where A is the standard and in the column where B is the standard. It would of course be possible to set up an experiment with a single standard initial stimulus which is itself made to increase or decrease gradually until the subject indicates that he has recognized the direction of change. Such a set-up would be unusual and it does not involve the phi-gamma hypothesis. It would be a variation of the limiting methods.

The *psychometric curve* is plotted with one of the stimuli chosen as a standard. The coördinates of this curve are as follows. The x-axis represents the variable stimulus magnitude. The y-axis represents the proportion of judgments "Variable stimulus greater than the standard stimulus." It is designated $p_{v>s}$. Note that the subject had nothing to do with the designation of any one or all of the stimuli as standards. He simply discriminated, as best he could, between pairs of stimuli.

When the variable stimulus is much smaller than the standard, then there will be very few judgments "$v > s$" because this judgment is incorrect and in the minority. Consequently $p_{v>s}$ will be small, *i.e.* $p_{v>s} < .50$. When the variable stimulus is equal to the standard, the judgments of the subject will be equally divided between "$v > s$" and "$v < s$" and hence $p_{v>s} = p_{v<s} = .50$. We are assuming that space and time errors have been eliminated or counterbalanced. When the variable stimulus is much greater than the standard, there will be a majority of correct judgments "$v > s$" so that $p_{v>s} > .50$.

The *phi-gamma hypothesis* states that the psychometric curve is the function described by the integral of the normal probability curve, namely

$$p_{v>s} = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{v} e^{\frac{-(v-s)^2}{2\sigma^2}}\,dv \qquad \text{(phi-gamma hypothesis).} \quad (1)$$

If $p_{v>s}$ is plotted against v we have the well-known ogive for the normal probability curve. The phi-gamma curve $\phi(\gamma)$ is symmetrical about its origin in the sense that the upper and lower quartile points of the ogive are equidistant from the mean. The phi-gamma hypothesis simply states that the psychometric curve can be fitted by the integral of the normal probability curve.

My first theorem merely asserts that if the absolute limen increases with the stimulus, then the psychometric curve for two categories of judgment is positively skewed. This theorem may be proved as follows.

Let there be n stimuli in the experimental series and assume, for simplicity, that they are equally spaced on the stimulus continuum. Let each one of the stimuli be compared with every other stimulus in the series. Let each pair be presented a sufficient number of times to stabilize the proportions $p_{v>s}$.

From such an experiment it is of course possible to tabulate the proportion of judgments "A is greater than B" for every possible pair of stimuli in the series. There will be $n(n-1)$ proportions in such a table. It is obvious that, since the subject is allowed only two categories of judgment, the proportion of judgments "A is greater than B" must necessarily be the complement of the proportion of judgments "B is greater than A." In other words,

$$p_{a>b} + p_{b>a} = 1. \quad (2)$$

Now let us collect all the proportions in which the stimulus R_2 is involved in order to plot the psychometric curve with that stimulus as a standard. It is represented in curve A of Figure 1. Let the experimental proportion

$$p_{1>2} = .25.$$

Then the lower quartile point for the curve A is R_1 and its median is R_2. Now, by the $\phi(\gamma)$ hypothesis, the curve is symmetrical and therefore the upper and lower quartile

points are equidistant from the median. But the stimuli are also equally spaced. Therefore the upper quartile point of curve A must be R_3. In other words,

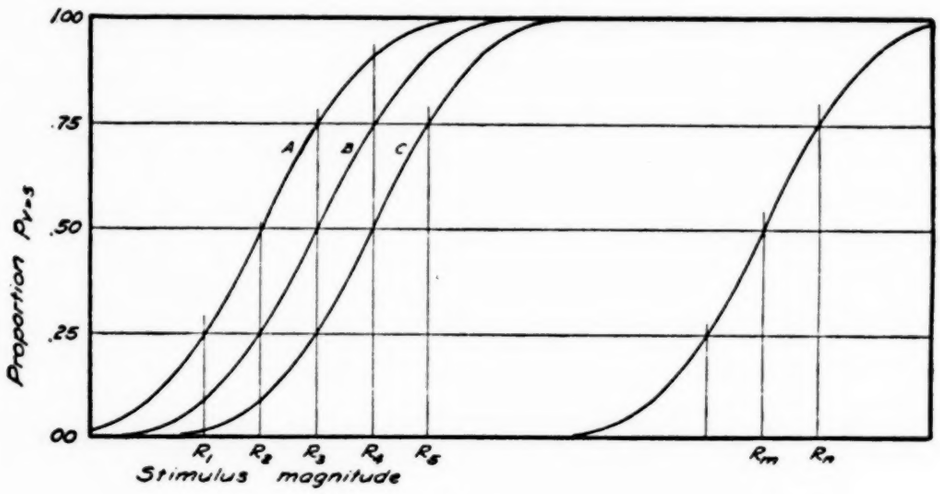

FIG. 1.

$$R_3 - R_2 = R_2 - R_1,$$

R_2 = median of curve A,

R_1 = lower quartile of A,

R_3 = upper quartile of A.

Therefore

$$p_{3>2} = .75.$$

But then, by (2)

$$p_{2>3} = .25,$$

and this locates the median and the lower quartile for curve B, in which R_3 is the standard. In other words,

$$R_4 - R_3 = R_3 - R_2,$$

R_3 = median of curve B,

R_2 = lower quartile of B,

R_4 = upper quartile of B.

This can be continued indefinitely so that

$$R_n - R_m = R_2 - R_1 \tag{3}$$

and

$$p_{n>m} = p_{2>1}, \tag{4}$$

where R_n is a large stimulus and R_2 is a small one.

But we have postulated that the absolute limen increases with the stimulus magnitude. Hence the two stimulus increments $(R_n - R_m)$ and $(R_2 - R_1)$ which are equally often correctly discriminated must be such that

$$(R_n - R_m) > (R_2 - R_1) \tag{5}$$

and hence the psychometric curve must be positively skewed rather than symmetrical. This is the first theorem. *It does not postulate Weber's law.* It merely postulates that the limen is an increasing function of the stimulus.

The first corollary is seen in Figure 1, in which the symmetry of the psychometric curve is postulated. It leads to equations 3 and 4 as the only condition under which the psychometric curve can be symmetrical.

The second corollary makes the necessary inference that if the absolute limen is an increasing function of the stimulus then the symmetrical phi-gamma curve cannot fit the psychometric curve which is positively skewed. This does not depend on Weber's law.

The third corollary introduces Weber's law. Weber's law states that the difficulty of discriminating between two stimuli is a function of the ratio of their magnitudes. The difficulty of discrimination is measured in terms of the relative frequency of correct discrimination. The law can be stated in algebraic notation as follows:[1]

$$p_{R>KR} = C. \tag{6}$$

If the constant C is arbitrarily set at .75, then the constant K represents the relative limen while R is the higher of the two compared stimuli. The relative limen constant is here less than unity. The same equation may be written

$$p_{KR>R} = 1 - C. \tag{7}$$

If it is desired to use the notation R' for the lower of the two stimuli, then the law takes the equivalent form

$$p_{R'l>R'} = C,$$

where

$$l = 1/k > 1,$$

[1] L. L. Thurstone, Three psychophysical laws, *Psychol. Rev.*, 1927, 34, 424–432.

and this can of course be turned into the complementary form

$$p_{R'1<R'} = 1 - C.$$

Weber's law asserts that the limen is proportional to the stimulus while the phi-gamma hypothesis requires that the limen be constant. Both cannot be true of the same set of data.

The second theorem is that if the absolute limen is proportional to the stimulus (Weber's law), then the positive skewness of the psychometric curve increases with the coarseness of the perceptual process. This theorem and its corollary regarding the necessary discrepancy between the symmetrical phi-gamma curve and the positively skewed psychometric curve can be shown perhaps best by several examples of different degrees of keenness of discrimination.

The psychometric curves in Figures 2, 3, and 4 are drawn

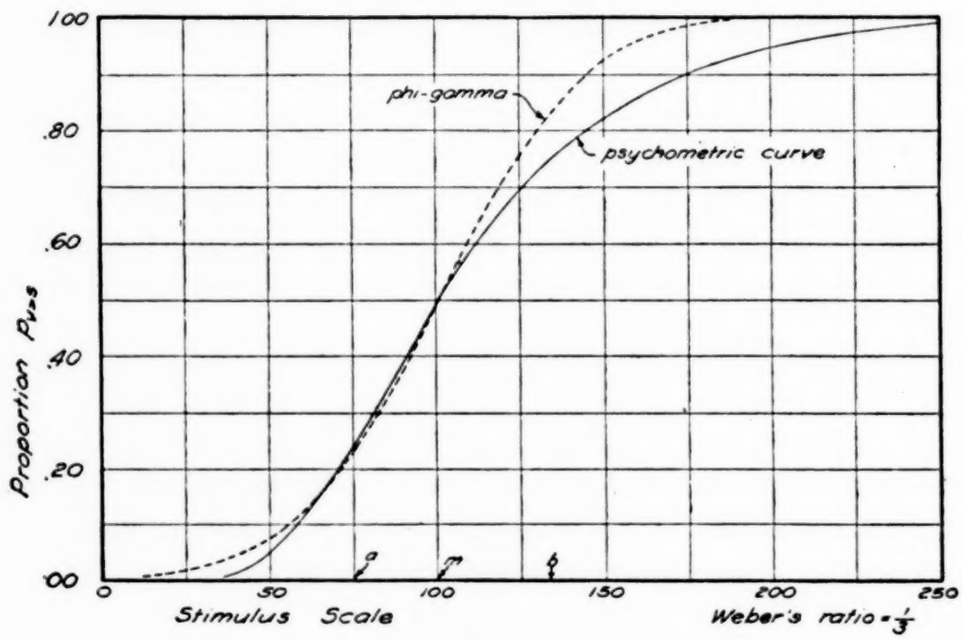

FIG. 2.

with the supposition that Weber's law holds in the vicinity of the magnitude of the standard stimulus. We do not need to assume that Weber's law holds for extremely large or for extremely small stimuli in order to draw these three psychometric curves. We shall designate the magnitude of the

standard stimulus artibrarily as 100 and we shall assume that Weber's law holds in the vicinity of this magnitude in order that a comparison may be made between the phi-gamma curve and the psychometric curve.

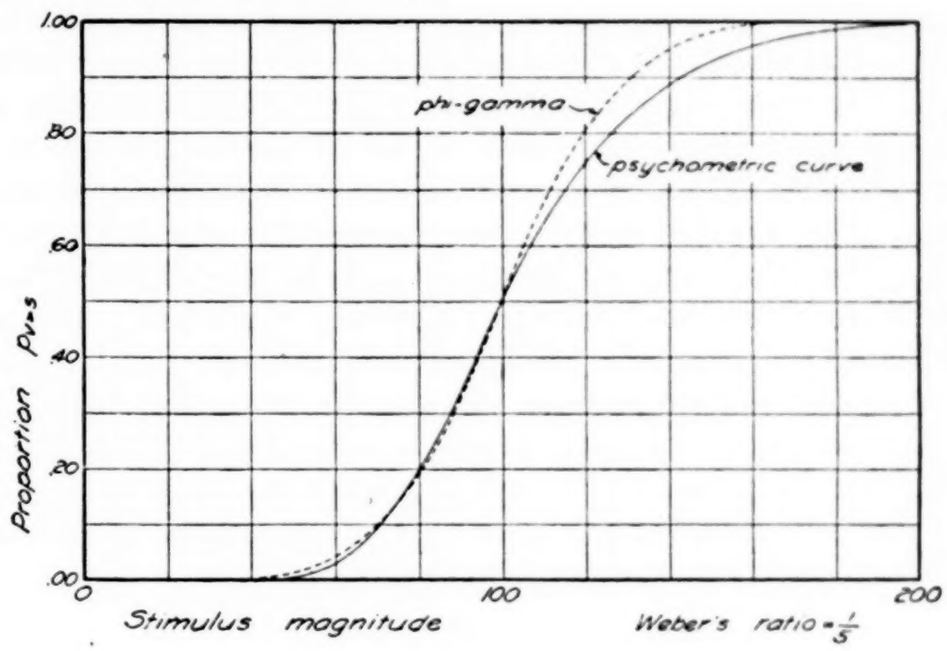

FIG. 3.

In Figure 2 we have taken Weber's ratio as 1/3 which represents coarse discrimination, such as for loudness of sound.

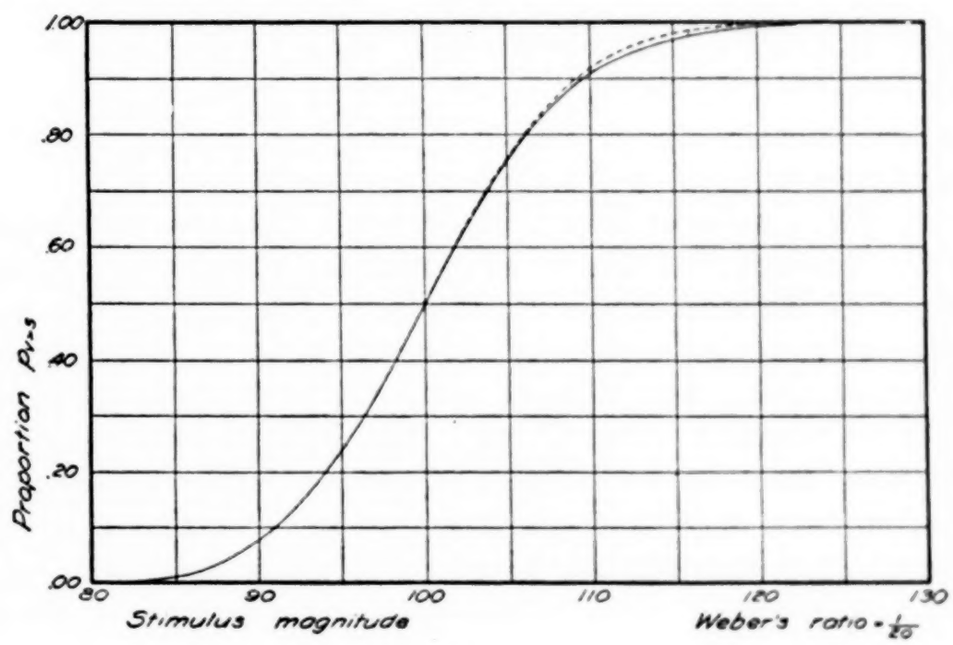

FIG. 4.

With the coarseness of discrimination represented by this ratio we shall expect two stimuli of magnitudes 100 and 133 to be correctly discriminated in 75% of the attempts. This is represented graphically by the fact that the psychometric curve of Fig. 2 reaches the level .75 when the variable stimulus is 133 and the standard stimulus is 100. Since Weber's ratio is here assumed to be 1/3, we shall expect the same frequency of correct discrimination, namely 75%, for two stimuli 75 and 100 since they have the same ratio as the two stimuli 100 and 133. The psychometric curve shows that the proportion $p_{75>100}$ has the value .25 so that $p_{100>75} = .75$. The psychometric curve has been drawn to represent the fact that it is as difficult to discriminate between two stimuli of 75 and 100 as it is to discriminate between two stimuli of 100 and 133 since these two pairs of stimuli have the same ratio. This is the condition described by Weber's law when the ratio is 1/3.

But it is immediately clear that this psychometric curve cannot possibly be symmetrical because

$$133 - 100 \neq 100 - 75$$

and consequently the curve cannot be Gaussian. The dotted line of Fig. 2 represents a phi-gamma curve which has been so drawn that it passes through the median and the lower quartile of the psychometric curve. If the phi-gamma curve must be fitted to the psychometric curve in this instance it can be moved to the right or left, and its standard deviation can be altered; but it cannot possibly be made to fit the psychometric curve which is clearly non-symmetrical.

In Fig. 3 we have drawn two curves in the same manner as in Fig. 2 except that here we have assumed Weber's ratio to be only 1/5. In other words, this figure represents a keener discriminatory function than the previous figure. Here, again, we are assuming that Weber's law holds in the vicinity of the standard stimulus value of 100. The psychometric curve is so drawn that the two stimuli of 100 and 120 are correctly discriminated in 75% of the attempts and also that the two stimuli 83.3 and 100 are as often correctly discriminated. The psychometric curve is here also non-symmetrical, since

$$120 - 100 \neq 100 - 83.3,$$

although the degree of assymetry is not so pronounced. The discrepancy between the phi-gamma curve and the psychometric curve is smaller than in the preceding figure where Weber's ratio was assumed to be larger.

In Fig. 4 we have the same type of diagram for a perceptual function in which Weber's ratio is much smaller, namely 1/20. In other words, when the difference between two stimuli is 5% of the smaller one, the frequency of correct discrimination is expected to be 75%. It will be seen in Fig. 4 that the discrepancy between the psychometric curve and the phi-gamma curve is relatively small since the discrepancy

$$105 - 100 \neq 100 - 95.2$$

is much smaller than in the previous examples.

Finally, when Weber's ratio is only one or two per cent, as in the perception of visual brightness, then the discrepancy is so small that the phi-gamma curve can be used to advantage in fitting the psychometric curve. It is of some interest, however, to know that the phi-gamma curve is theoretically incorrect, that it cannot possibly be the correct function for the psychometric curve, and that its usefulness is merely that of an empirical equation or curve. It serves the purposes of a scroll, as it were, in drawing a smooth curve for the psychometric data with an error that is very small, perhaps smaller than the observational errors, when the perceptual function which is being studied is keen.

The fact that the fallacious character of the phi-gamma hypothesis has not previously caused trouble is to be found in the fact that most of the perceptual functions that have been investigated have been rather keen. Examples of these are the perception of visual brightness and of line-lengths. In cases of coarse discrimination such as that of sound intensity there has been the uncertainty of stimulus measurement which introduces another ambiguity into the handling of psychometric data. When the perceptual function is rather coarse, it is necessary to take into consideration the error of the phi-gamma hypothesis.

The third theorem is that if Weber's law is satisfied by the

psychometric data, then a $\phi(\log \gamma)$ hypothesis will also be satisfied by the same data. If we want an empirical equation for the psychometric function which shall be at least not theoretically inconsistent with Weber's law, then we may propose as a substitute for the $\phi(\gamma)$ hypothesis a $\phi(\log \gamma)$ hypothesis. The integral of the normal curve would then be fitted on a psychometric function in which the proportions $p_{v>s}$ are plotted against the logarithm of the stimulus intensity.

Even if Weber's law is not exactly verified by the data it is in the great majority of experiments more nearly true to assume that the limen is proportional to the stimulus than to say that the limen is constant and independent of the stimulus magnitude. The latter assumption of the phi-gamma hypothesis flies in the face of the simplest experimental evidence.

Let the points a and b of Figure 2 be the base line values corresponding to the ordinates .25 and .75 respectively of the psychometric function. If the empirical equation used in fitting the psychometric function is to be symmetrical, then

$$b - m = m - a. \tag{8}$$

Let

$$a = \log R_a,$$
$$b = \log R_b,$$
$$m = \log R_m.$$

Then, by (8)

$$\log R_b - \log R_m = \log R_m - \log R_a$$

or

$$\log \frac{R_b}{R_m} = \log \frac{R_m}{R_a}$$

and hence

$$\frac{R_b}{R_m} = \frac{R_m}{R_a},$$

which is the condition required by Weber's law.

This manner of fitting the psychometric function is at least consistent with Weber's law and it may be called the $\phi(\log \gamma)$ hypothesis as a substitute for $\phi(\gamma)$ hypothesis. It should be regarded merely as an empirical equation for determining the limen and the precision. It should not be taken too seriously either in the form $\phi(\gamma)$ or in the form $\phi(\log \gamma)$.

The actual fitting of the curve may be accomplished graphically to a fair degree of accuracy by means of arithmetic-probability cross-section paper, or it may be done by plotting the sigma value corresponding to the experimental proportion against the logarithm of the stimulus magnitude. The plot should be linear. A constant error will be revealed by the intercept, and the slope of the linear plot will be the precision. The function can also be fitted by the method described by Kelley,[2] with this difference; the logarithm of the stimulus should be used instead of the stimulus magnitude. The system of weighting which is ordinarily used need not be modified because the weight is a function of the observed proportions which are retained as before. Urban's tables and Rich's checking tables are unfortunately not ordinarily applicable to this modification of the phi-gamma hypothesis because these tables presuppose that the stimulus magnitudes are equidistant. They would be applicable if the experiments were so set up that the logarithms of the stimulus magnitudes should constitute an arithmetical series.

It should be noted that my main criticism of the phi-gamma hypothesis is not contingent on the acceptance of Weber's law. The phi-gamma hypothesis is denied if the absolute limen increases with the stimulus magnitude even if Weber's law is not exactly verified.

(Manuscript received Jan. 5, 1928)

[2] T. L. Kelley, *Statistical method*, 1924, 326–328.

THE RÔLE OF ARTICULATION IN MEMORIZING

BY M. C. BARLOW

University of Utah

Economic procedure in learning, in so far as it touches the use of articulation, can be studied by way of memorization. This paper attempts to find the advantage, if any, possessed by explicit laryngeal responses over implicit, or non-laryngeal, responses in memorization.

The subject-matter consisted of two 20-word lists of three-letter syllables that conveyed no obvious meaning. Each syllable contained a vowel between two consonants. They were printed upon strips of paper in such a way that they could be exposed singly for a given period of time. The exposures were made by means of a frame-like apparatus of wood with two rollers vertical to one another.

Subjects to the number of 327 memorized the learning material. These pupils range in school from the second to the ninth grade, in addition to 51 students of college grade. Each grade took the test in two groups, so as to learn the word-lists in balanced order. Half of the class practiced first the A-list or vocalization-list, then the B- or non-vocalization list. The other half learned the lists in the reverse order. This procedure distributes practice gains about equally between the two lists.

The instructions were approximately as follows. "I am going to show you a list of words one at a time. You are to learn as many of these words as you can. As an aid in learning them you are to repeat the words over and over to yourself as you study them. Use your tongue and lips as you say them in whispering tones. When you have seen the words four times each, and when I say 'write,' you are to write as many of the words as you can remember. If you are not sure of the word, guess."

The experimenter exposed each syllable four times, allowing three to four seconds for each syllable.

The instructions for the second list were; "Now I am going to show you a second list of new words. You are also to learn as many of these words as you can. This time you are to practice without saying the words over to yourself. To help you in this you are to hold your pencil between your teeth like this (experimenter demonstrated what was wanted by placing and holding a pencil crosswise between his teeth, the tongue being unrestrained)." For pupils who learned the B-list first, instructions were given which differed from the foregoing only in such a way as to take into consideration the necessary difference in procedure.

The subjects were allowed five minutes for the reproduction of learned syllables. The score was taken as the number of words correctly reproduced.

The gross scores appear in the A and B columns of Table I. The mean gross scores from the verbal method are shown by *A* and those of the non-verbal by *B* in Fig. 1. The numerical differences between the scores made from the two methods of

learning are found to be larger for the higher than for the lower grades. The smaller differences in the lower grades are obviously due to their lower memorizing ability.

TABLE I

TWO 20-SYLLABLE LISTS LEARNED BY METHODS OF 'VERBALIZATION' AND 'NON-VERBALIZATION'

A = average number syllables reproduced by verbalization method; *B* = the same with instruction against articulation; PE(diff) = probable error of the difference *A* − *B*; *D*/PE(diff) = PE(diff) units in the difference of the means *A* − *B*; *P* = chances in 100 that the real difference of means *A* and *B* exceeds 0; *Ri* = total number right syllables (*A* + *B*); *Wr* = total syllables wrong; *N* = number of cases.

Grade	*A*	*B*	PE(diff)	*D*/PE(diff)	*P*	*Ri*	*Wr*	*N*
2	2.27	2.00	.440	.61	66	4.27	6.82	22
3	5.04	4.03	.536	1.88	89	9.07	6.53	34
4	6.43	6.11	.648	.49	63	12.54	9.00	37
5	8.50	7.44	.624	1.54	85	15.94	8.62	40
6	8.42	7.39	.803	1.28	81	15.81	9.53	36
7	9.63	8.13	.686	2.18	93	17.76	10.06	37
8	10.90	9.58	.840	1.58	86	20.48	8.02	40
9	11.78	10.59	.880	1.35	82	22.37	6.72	32
College	15.34	12.42	.757	3.86	99.5	27.76	4.69	49
All groups	9.37	8.16	.333	3.64	99	17.30	7.03	327

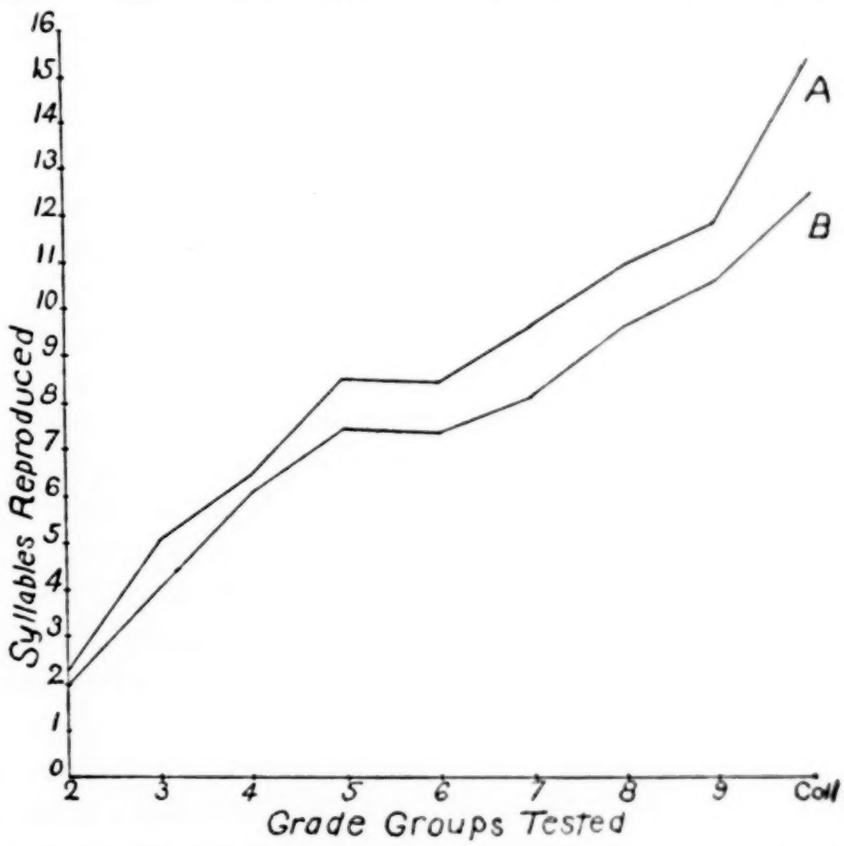

FIG. 1. *A* = number of syllables reproduced from 20-word lists learned with verbal responses. *B* = number of syllables reproduced with restricted articulation.

It is necessary to know the reliability of these differences. What are the chances that differences will continue to be found with an extension of the study to include more subjects? The probable error difference formula $PE(diff) = \sqrt{(PE_{M1})^2 + (PE_{M2})^2}$ serves our purpose here. With twenty-two cases in the Second Grade, the chances are 66 to 100 favoring superiority of the vocalization method of learning; 89 chances of 100 for the third grade; 63 for the fourth; 85 for the fifth; 81 for the sixth; 93 for the seventh; 86 for the eighth; 82 for the ninth; and 99.5 for the college group. Further, the chances are 99 out of 100 that the better scores made by the use of language in memorizing, taking the group as a whole, represent a true difference and not a difference due to chance conditions.

Sex differences are found to favor the girls. Girls excel the boys in memorizing. The mean number of syllables learned by 153 males is 8.79, the mean score of 173 girls is 9.73. The boys' scores have a wider range than the girls'. The sigma of the distribution is 4.73 for the boys and 4.39 for the girls. The PE(diff) of the mean scores is .695. There are 1.35 PE(diff) units in the difference of the mean score of girls and boys. This means that in 82 per cent. of the cases the girls excel the boys.

Again the girls excel the boys in memorizing with restricted articulation. For the above number of boys the mean number of syllables learned by the 'non-language' procedure is 7.55, with a sigma of the distribution of 4.18. The mean for the girls is 8.51 with a sigma of 4.18. The PE(diff) is .66. Hence the chances are that the girls will exceed the boys 84 times in a 100 for learning syllables by the 'non-language' method.

Individuals show wide differences in favor of one or the other method of learning. Of the 327 cases, 52 subjects reproduced the same number of syllables by both methods; 34 pupils reproduced one syllable more by Method A; 45, two syllables more; 35, three syllables; 29, four syllables; 13, five; 10, six; 12, seven; 3, eight; 2, nine; 1, eleven; and 1, twelve. There were 33 pupils who learned one syllable more by Method B; 30 cases learned two more by this method; 11, three more; 7, four; 2, five; 3, six; 3, seven; and 1, nine. 185 pupils learned more words by the former and 90 learned more words by the latter method. Although substantial differences are found in favor of the 'vocalization' method, there is no justification for claiming that all children should memorize material by this means. In the light of the large amount of overlapping of the scores obtained by the use of the two methods, it is probably safe to conclude that the best practical results in memorization can be secured by having the individual learners follow the method which yields best results. Experimental procedure appears to be the more satisfactory way of determining the preferred method of memorization, for the pupil is usually in no position to know beforehand which method is really better for him.

The number of wrong syllables produced is another difference factor for consideration. The greatest amount of overlapping from the two methods appears among the scores in grades five to eight, while in the second, third and fourth grades there is less of it. There is also less overlapping of errors made from the two methods as they apply to the ninth grade and also to the college group.

Fewer errors are made with restricted articulation. This fact seems to stand in favor of this method of learning as against the other. There are, however, qualifying factors which deserve to be considered. It has already been shown that because language is beneficial in learning its use results in the recall of a larger amount of learning material. In the lower grades, particularly the second, third and fourth, pupils who make more wrong responses also reproduce more right syllables.

In the upper grades the foregoing condition is reversed; *i.e.* high scores occur with a small number of errors. The small number of errors made by pupils of the higher grades is in all probability due to abridgments and abbreviations in the learning process.

Since more errors result from the non-articular method in the eighth grade and above, as is found in this study, it is quite fair to assume that for all groups above the eighth grade the errors will be fewer for the verbal method. Though there are no data here on the tenth, eleventh and twelfth grades, the relatively smaller number of errors for the grades we have used seems to justify the above statement.

These results suggest that, other things being equal, memorization is not only more accurate but also easier with articulation than without, for pupils in the primary as well as the upper grades. It is quite probable that intermediate grades also would yield similar results but for the influence of extensive training in silent reading in the primary grades. If the silent method is more natural, which seems not to be the case with our subjects, pupils that are once trained to read and think without articulation should go on using the same procedure more economically after the demands for silent reading have become less exacting.

The syllables produced by the second grade (Table I and Fig. 2, *Wr*) are more often wrong than right. From a knowledge of this condition it is not surprising that the reportability of children is very unreliable. It is also of interest to note that third grade children make more right than wrong responses. The total number of errors (*Wr*) shows a gradual increase up to the seventh grade, with a steady decline from here up.

The total number of right responses (Table I and Fig. 2, *Ri*) made by these groups presents a different picture. When the two variables are plotted, *i.e.* the total number of right responses for both lists and the successive grade groups, the graph is approximately linear. It is quite probable, however, that if the results from the ninth, tenth, and eleventh grades were included, the graph would show the region of diminishing returns so characteristic of learning curves.

Since we gave only one form of task, there are no scores from which to calculate reliability measures. There is, however, reason for imputing reliability to the scores when we consider their differences all together. In the nine groups there were over 300 subjects, and every group yielded scores favoring the same method of learning. Considerable claim to high reliability seems reasonable when a number of groups of subjects made scores with differences in the same direction, even where those differences are small.

A word concerning the possible objection that pencils held in the mouths of pupils are detrimental to non-language

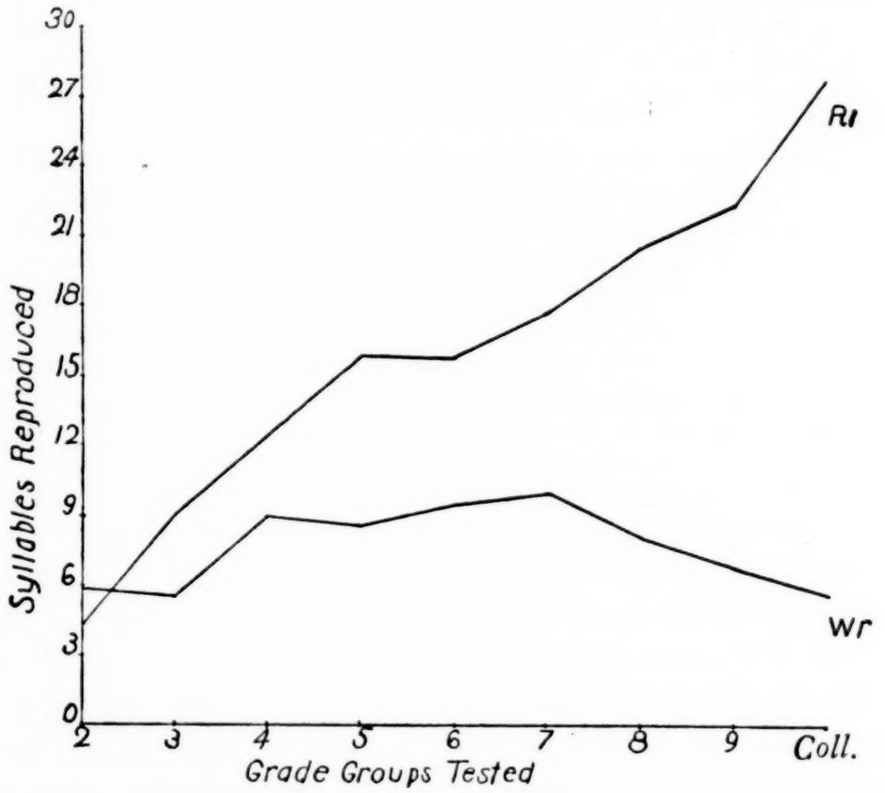

FIG. 2. *Ri* = average number of syllables correctly reproduced by various groups from two 20-syllable lists one learned with verbalization and the other with restricted articulation. *Wr* = average number of wrong responses from the same lists.

learning and beneficial to language learning. If this be true, it is possible that we have here a spurious factor not actually related to either method of learning. The writer observed few pupils whose behavior indicated that the pencils were real distractions. Some subjects manipulated the pencils while practicing. Others reported difficulty in learning because of these distractions. One pupil was heard to say that he repeated the words subvocally, with the pencil between

his teeth. Other reports indicated that the tendencies to vocalize were felt even with the silent learning. It is probably not far from the truth to say that the inhibitory influences caused by the pencils in the case of non-vocalization about equal the facilitating influences of vocalization in the case of the inarticulate procedure.[1]

Our grade school subjects came from the University of Utah Training School where they are required to read silently, that is, without sub-vocalizing the words and without lip movement. It is quite likely that the previous training of any pupil exerts a measurable influence upon his ability to memorize by either of the foregoing methods. Since our subjects were practiced in silent reading, we should expect this kind of training to favor higher scores from silent than from verbal memorization, but the results are contrary to our expectations. We regard this as a tentative conclusion, only, pending the outcome of further experimentation with other controls of articulation than those already employed. These results indicate that the problem of linguistic responses is of considerable importance within the field of memorization.

How can we account, in terms of theory, for the advantage in memorization which is seemingly due to articulation? A plausible explanation appears to be in terms of the dynamic influence of articulation upon learning mechanisms. Each word, as it is exposed on the memory drum, is a stimulating condition which is antecedent to, and partially determinative of, the consequent language response. The stimulus may be said to evoke the response. In accordance with the theory of the reflex circuit, this language response now becomes a component part of a new stimulating situation in such a way as to constitute a dynamic factor in learning. Repetitions of the language-word act facilitate the fixation of right responses in the neuro-muscular equipment concerned with learning. Considered more broadly, it may be said that every explicit language response is a concrete fact of mental achievement within the experience of the individual. It is evident that more attention should be given to the study of our responses considered as stimuli to further responses.

It may be implied that memorization of meaningful material can be accomplished more economically through vocalization. To establish this point beyond question it is desirable to test the learning of meaningful material by a technique similar to that used in this experiment.

Summary

(1) The average result from every one of our groups shows that the syllables were more economically learned with full than with restricted articulation. (2) The girls make better scores than the boys in both kinds of learning. (3) The pronounced individual differences in the relative number of syllables learned by the two methods point to a need of experimental determination for each pupil of the better

[1] *Editor's note.* The discriminating reader will distinguish between *non-verbalization* and *restricted articulation.* The presence or absence of verbal symbols is not controlled in the present study.

method for his use in memorization. (4) The errors made by the first-described method are slightly in excess of those made by the latter. They are about equally numerous in the lower and in the upper grades, and more numerous in the intermediate grades.

Further related problems of immediate interest include (1) The influence of verbalized learning upon retention of the memorized material; (2) The rôle of language in learning to spell, and (3) The relation between the degree of preference for either method of learning and general intelligence.

(Manuscript received Dec. 23, 1927)

THE PREDICTION FROM PHOTOGRAPHS OF SUCCESS AND OF VOCATIONAL APTITUDE

BY CARNEY LANDIS AND L. W. PHELPS

Wesleyan University

A general and popular belief persists that vocational aptitude and success can be judged from photographs.

The widespread belief in this system of character analysis is illustrated by the popularity of the 'analyst' Katherine Blackford, whose 'character analyses' are based in part upon physiognomy (2). The elements of personal fitness for a vocation are for her six: health, character, intelligence, disposition to industry, natural aptitudes and experience. These six she considers as of equal importance. She classifies men as *types* according to their natural endowments. Each type is by nature definitely fitted for one general sort of vocation and unless the type and the vocation coincide an industrial misfit is supposed to result. In fact, if men are so unfortunate as to get into life-work which does not allow them to expand their physical and mental energies in a normal way, they are likely to descend to crime; they increase labor turnover and they represent a definite economic waste. Blackford makes her analyses of character on the basis of a photograph or of a personal interview.

This belief that we are able to 'size a man up,' to judge his character, to tell something of his personal qualities, is widely prevalent. Witness the usual request for a photograph with an application for a position, the adding of photographs to the personnel folder, and so on.

Pintner (4) published an interesting study several years ago in which photographs of twelve children were presented to five classes of observers, who were asked to rank the photographs in order of 'intelligence.' The judges were nine physicians, fifteen psychologists, seventeen teachers, eleven students and eleven unclassified individuals.

The correlations of the judgments from the photographs to 'intelligence' as determined by tests were so low that Pintner concluded that intelligence cannot be judged from photographs. The highest single coefficient, + .52, was made by one of the miscellaneous group, but a − .51 also appeared in this same group. The coefficients for the psychologists ranged from + .45 to − .63. The median coefficient of correlation for all observers was + .10. The chance arrangements of the photographs correlated with the rankings according to the mental tests almost as well as the rankings of the observers. Pintner concluded that a five-minute interview with the children would have enabled the observers to make more accurate judgments.

Anderson (1) made a similar study in which he used sixty-nine photographs of the employees of a large company. The twelve judges were graduate students and instructors in psychology. The criterion of actual intelligence was an adaptation by the Bureau of Personnel Research of the Army Intelligence Tests. The observers grouped the subjects by an arrangement which sifted out the extremes and left out an undifferentiated middle group of about 40 per cent. From his study Anderson concluded that it is very doubtful whether intelligence can be gauged from

a photograph well enough for any practical purpose. The fact that there was a correlation of + .27 between the assigned ratings and those obtained from the criterion indicates, however, that, in this instance at least, there was a tendency to discriminate intelligence above chance selection. The impracticability of using the method of the observation of photographs in the employment office is shown by the fact that over and over again the same subjects were assigned ratings both in the group composed of those of superior intelligence and in that composed of those of inferior intelligence.

A recent study by Gaskill, Fenton and Porter (3) reports results differing from those reported by Pintner and by Anderson. It was based upon photographs of twelve boys of the same age. The I.Q. range of this group of boys was from 18 to 171. These photographs were shown to 274 observers who ranked the photographs on the basis of the intelligence which they judged each child to possess. The median correlation between the actual I.Q. and the estimated intelligence was + .42. (It may be noted that Pintner (4) reports correlations as high as + .43 obtained by pure chance arrangements.)

It seems to us that Pintner and Gaskill, together with several other experimenters who have used this method, have overlooked the fairly obvious point that it would be possible so to select the photographs that the bright children looked 'bright' and the dull ones looked 'dull,' or *vice versa*. We do not mean to suggest that any experimenter has so selected his photographs, but we point out that the subjects whose photographs are to be used in an experiment of this type should be selected on the basis of their I.Q. record, sales-ability, *etc.*, and then, after the selections have been made, photographs should be obtained.

These studies which have been made of the ability of observers to judge intelligence on the basis of photographs are the most direct attack thus far made on the sweeping claims of the physiognomists. The present report offers further experimental evidence in contradiction to their claims and assumptions. Our purpose was to attempt to find answers to the following questions.

(1) Can untrained observers predict probable vocational success or failure on the basis of a photograph?

(2) From an examination of his photograph, can observers determine whether a man who has been following one vocation for 25 years has been relatively successful or unsuccessful in it?

(3) Is it possible to predict vocational aptitude from photographs?

(4) What is the basis of judgment of vocational aptitude when the judgment is made from a photograph?

(5) How certain are observers of judgments of this character?

METHOD

An alumni publication of a large American university was secured. This publication was issued on the 25th anniversary of the graduation of a class of 850 men and it contained an autobiographical sketch of each man together with his photograph at the time of graduation and at his 25th anniversary of graduation. For the purposes of this study, all autobiographies of men who had gone into law, medicine, education and engineering were read and assigned a rating from 1 to 10 on the basis of their apparent worldly success. For instance, a lawyer who had risen to be one of the chief attorneys for the Standard Oil Company was judged relatively successful, while one who was acting as a clerk in a large law office was relatively unsuccessful. A man who had taught for 25 years in the common schools was judged relatively unsuccessful, while one who had risen to the superintendency of schools in a large city was judged relatively successful. In this study it must be borne in mind that success or non-success relates only to the vocation and is always a relative term—relative to the performance of other members of his own college class.

From the ratings thus assigned the five most successful and the five least successful in each of the four groups, law, medicine, education and engineering, were taken. The published portraits were re-photographed and mounted on separate cards.

We feel that this method of selection meets the criticism which we urged against other investigations; namely, the possibility of using photographs which portray the traditional facies of the trait in question. Our selection was not made on the basis of the photograph, but on the basis of the autobiography. Hence, whether the man 'looked' successful or otherwise had nothing to do with his inclusion in the group of selected men.

These photographs were presented, one at a time, to 20 observers, undergraduates at Wesleyan University, all of whom had had at least an elementary course in psychology. Each observer was given two prepared blanks on which to record his judgments. He was not told that the two groups of pictures were of the same men. The first blank used by the observers bore the following instructions:

"This experiment is designed to test experimentally your ability to predict vocational success in life on the basis of photographs of college seniors; also to test your ability to predict what profession, vocation, craft, or trade these men will follow. The photographs will be presented to you singly. Look at each and then record your judgments in the appropriate space below. Under 'Success or failure' place an S or F. S signifies that you believe the man will be relatively successful in whatever vocation he may follow. In the next column record the "Per cent. certain" of this judgment, 0 per cent. meaning a hunch or no feeling of certainty whatever, 50 per cent. moderately certain and 100 per cent. meaning positive conviction. In the column headed 'Vocation' indicate the vocation, profession, trade, etc., which you would recommend. In the column labeled 'Basis of Judgment' enter the reason best describing your basis of judgment."

The columns on this blank were headed (1) picture number, (2) vocation, (3) success or failure, (4) per cent. certain, (5) basis of judgment, (6) remarks.

Blank 2 carried the following instructions:

"This experiment is designed to test experimentally your ability to determine what vocation, profession or trade men have followed for 25 years and what relative success they have attained in their vocation.

"The pictures will be presented to you singly. Look at each and then record your judgments in the appropriate space below.

"Under the column labeled 'Vocation' indicate the vocation which you believe the man has followed the past 25 years. In the column headed 'Success or failure' put an S or F indicating that you believe that the man has been *relatively* successful or *relatively* unsuccessful. In the column headed 'Per cent. certain' record your certainty, 0 per cent. indicating no feeling of certainty whatever, 50 per cent. moderately certain and 100 per cent. positive convictions. In the column labeled 'Basis of judgment' enter the reason best describing your basis of judgment."

The columns on this blank bore the same headings as Blank 1.

Results

In Table I the numbers from S1–O to S20–O indicate photographs of the men who had been successfully following their vocation for twenty-five years. An examination of the

Table I

Most Successful						Least Successful					
Picture	No. S Jdgmts	No. F Jdgmts	% S Jdgmts	% F Jdgmts	Av. % Cert'y	Picture	No. S Jdgmts	No. F Jdgmts	% S Jdgmts	% F Jdgmts	Av. % Cert'y
S1–O	18	2	90	10	40.7	F21–O	12	8	60	40	36.7
S2–O	19	1	95	5	39.7	F22–O	17	3	85	15	44.4
S3–O	10	10	50	50	28.5	F23–O	18	2	90	10	32.3
S4–O	13	7	65	35	35.5	F24–O	11	9	55	45	32.0
S5–O	16	4	80	20	30.7	F25–O	15	5	75	25	28.0
S6–O	15	5	75	25	40.0	F26–O	6	14	30	70	28.3
S7–O	14	6	70	30	36.5	F27–O	12	8	60	40	41.2
S8–O	11	9	55	45	28.2	F28–O	14	6	70	30	32.7
S9–O	14	6	70	30	31.3	F29–O	10	10	50	50	35.2
S10–O	10	10	50	50	34.2	F30–O	10	10	50	50	38.3
S11–O	15	5	75	25	30.0	F31–O	12	8	60	40	39.7
S12–O	18	2	90	10	24.1	F32–O	12	8	60	40	34.4
S13–O	12	8	60	40	33.6	F33–O	18	2	90	10	42.0
S14–O	9	11	45	55	36.3	F34–O	15	5	75	25	36.3
S15–O	18	2	90	10	29.2	F35–O	12	8	60	40	31.9
S16–O	7	13	35	65	26.7	F36–O	14	6	70	30	35.0
S17–O	18	2	90	10	38.1	F37–O	17	3	85	15	43.6
S18–O	9	11	45	55	34.8	F38–O	10	10	50	50	34.7
S19–O	9	11	45	55	29.7	F39–O	15	5	75	25	36.5
S20–O	18	2	90	10	36.7	F40–O	10	10	50	50	27.7
Av.	13.7	6.3	68.3	31.7	34.2	Av.	13.0	7.0	65.0	35.0	35.5

results of the judgments as to whether these men were or were not successful shows that in only four cases did more than 50 per cent. of the observers think that the subject belonged to the 'least successful' group. In the case of picture S14–O, 55 per cent. of the observers placed the subject in the least successful group, the man represented in picture S16–O was judged relatively unsuccessful by 65 per cent. of the observers, S18–O by 55 per cent. of the observers, and S19–O by 55

per cent. of the observers. In truth, all of these four belonged to the 'most successful' group. Number S14–O is an extremely successful lawyer and numbers S16–O, S18–O and S19–O are all outstanding men in the medical world. The fact that group-judgment pronounced fourteen of the twenty men successful (in two cases the group judgments were equally divided), might seem to be indicative of the presence of some determining factor.

When we examine the judgments made regarding pictures F21–O through F40–O, which consisted of the men who were the 'least successful,' it seems that the relatively accurate judgments of the observers concerning pictures S1–O through S20–O may have been due to the optimism of those making the judgments regarding success more than to the perception of some determining factor of actual ability. The group-judgments of the pictures of the 'least successful' men in only one case nominated a man as relatively unsuccessful. This was subject F26–O who was judged, on the basis of his picture, 'S' by six and 'F' by fourteen observers. The judgments were equally divided in the cases of numbers F29–O, F30–O, F38–O and F40–O, 50 per cent. judging the subjects successful and 50 per cent. judging them relatively unsuccessful.

Table II shows substantially the same situation as Table I. Numbers F21–Y through F40–Y represent the group of college seniors who twenty-five years later composed the 'least successful' group in this study. Numbers F24–Y, F33–Y, F34–Y and F36–Y were the only subjects in this group who were assigned to positions among the least successful by a majority of the observers. In the case of picture F27–Y the judgments were equally divided.

On the other hand, only two of the subjects who actually belonged to the 'most successful' group (numbers S1–Y through S20–Y) were placed in the 'least successful' group by a majority of the observers.

The average percentage of certainty of the entire group of observers for their judgments of the success of the entire group of older men is 34.8 per cent.; for the group of most

successful older men, 34.2 per cent.; for the group of least successful older men, 35.5 per cent. For the group of college

TABLE II

Most Successful						Least Successful					
Picture	No. S Jdgmts	No. F Jdgmts	% S Jdgmts	% F Jdgmts	Av. % Cert'y	Picture	No. S Jdgmts	No. F Jdgmts	% S Jdgmts	% F Jdgmts	Av. % Cert'y
S1-Y...15	5	75	25	20.7	F21-Y...15	5	75	25	28.2		
S2-Y...13	7	65	35	23.0	F22-Y...13	7	65	35	34.5		
S3-Y...13	7	65	35	32.7	F23-Y...11	9	55	45	34.8		
S4-Y...14	6	70	30	30.7	F24-Y... 8	12	40	60	25.0		
S5-Y...17	3	85	15	30.0	F25-Y...15	5	75	25	30.5		
S6-Y...11	9	55	45	33.0	F26-Y...17	3	85	15	31.1		
S7-Y... 7	13	35	65	33.8	F27-Y...10	10	50	50	25.2		
S8-Y...13	7	65	35	25.5	F28-Y...15	5	75	25	31.2		
S9-Y...17	3	85	15	32.4	F29-Y...14	6	70	30	25.5		
S10-Y...16	4	80	20	30.7	F30-Y...13	7	65	35	28.7		
S11-Y...16	4	80	20	32.2	F31-Y...13	7	65	35	26.7		
S12-Y...16	4	80	20	32.0	F32-Y...16	4	80	20	32.0		
S13-Y...15	5	75	25	33.0	F33-Y... 9	11	45	55	30.4		
S14-Y...17	3	85	15	33.8	F34-Y... 9	11	45	55	30.0		
S15-Y...13	7	65	35	30.9	F35-Y...17	3	85	15	26.3		
S16-Y...18	2	90	10	25.5	F36-Y... 8	12	40	60	26.6		
S17-Y... 7	13	35	65	25.0	F37-Y...19	1	95	5	32.0		
S18-Y...12	8	60	40	32.0	F38-Y...14	6	70	30	35.2		
S19-Y...12	8	60	40	29.5	F39-Y...16	4	80	20	29.5		
S20-Y...16	4	80	20	20.0	F40-Y...12	8	60	40	22.9		
Av.....13.9	6.1	69.5	30.5	29.3	Av.....13.2	6.8	66.0	34.0	29.3		

seniors the average percentage of certainty of the observers with regard to success or failure for the group which turned out to be most successful is 29.3 per cent.; for the least successful group, 29.3 per cent.; for the entire group of the young men, 29.3 per cent.; for all subjects, in both groups, 32 per cent.

We found very little correlation between actual vocations and the vocations assigned to the subjects by the observers (Table III). In practically every case the observers disagreed and the same subject might be assigned to from ten to fifteen different vocations. We have selected the cases in which five or more observers placed the same subject in the same one of the fifty vocations.

With regard to the basis of judgment an examination of Table IV shows that the most common reasons for terming one of the older men either successful or relatively unsuccessful are general impression, similarity to some acquaintance, facial

TABLE III

OLDER MEN

Picture	Assigned Occupation	Jdg-mts	No. S	No. F	Actual Occup'n	Actual Standing
S10–O	Clergyman	10	6	4	Engineer	S
F23–O	College instructor	6	6	0	Educator	F
F25–O	Banking	6	6	0	Educator	F
F31–O	Retail Merchandising	5	2	3	Law	F

COLLEGE SENIORS

F22–Y	Clergyman	7	3	4	Medicine	F
F26–Y	Clergyman	5	5	0	Law	F
F31–Y	Clergyman	6	3	3	Engineer	F
F35–Y	Clergyman	5	4	1	Engineer	F
S8–Y	Clergyman	5	2	3	Engineer	S
S9–Y	Gentleman of leisure	6	4	2	Engineer	S
S14–Y	Law-gen'l practice	6	5	1	Law	S
S15–Y	College instruction	6	6	0	Law	S
S18–Y	College instruction	5	4	1	Medicine	S
S19–Y	Musician	6	4	2	Medicine	S

expression, or 'hunch.' The reasons occurring with least frequency are shape or size of ears and position of head.

TABLE IV

Basis of Judgment	OLDER MEN *Frequency*	*%*	COLLEGE SENIORS *Frequency*	*%*
Shape of Face	28	3.5	25	3.1
Appearance of Eyes	57	7.1	88	10.1
Length of Nose	6	0.75	7	0.9
Shape or Size of Mouth	38	4.8	37	4.8
Width of Forehead	6	0.75	7	0.9
Height of Forehead	9	1.1	13	1.6
Shape or Size of Ears	3	0.37	1	0.12
Condition (Neatness, etc.)	24	3.0	12	1.5
Facial Expression	103	12.9	117	14.6
Position of Head	4	0.5	4	0.5
Similarity to Some Acquaintance	112	14.0	122	15.3
Observation of Persons of similar Appearance	78	9.8	53	6.6
General Impression	200	25.0	179	22.4
Hunch	83	10.4	70	8.8
Snap Judgment	39	4.9	46	5.7
Unclassified	10	1.25	19	2.4

The bases of judgment having the greatest frequency in the case of the college seniors were the same as for the older men except that 'appearance of eyes' is substituted for 'hunch' in the fourth place in the former group. The same two bases appeared least frequently here, *i.e.* shape or size of ears and position of head.

An additional exercise was carried out as a check upon certain of the above results. The photographs were projected upon a screen by means of a projection lantern before classes of from 20 to 30 undergraduates. These students were given oral instructions as follows:

"I will project on the screen the photographs of 40 young men and of 40 older men. The photographs of young men were taken when the men were seniors in college some 30 years ago. The older men had been graduated from college some 25 years at the time the photograph was taken. Certain of these pictures are of men who will be or have been very successful in the profession which they follow. Other of the pictures are of men who will be or have been relatively unsuccessful in their professions. Note then that there are two groups, one of very successful men and one of relatively unsuccessful. Judge each picture in turn as S or F depending on whether you think the man has been or will be very successful or relatively unsuccessful."

The pictures were shown to 93 untrained judges. The results are given in Table V.

TABLE V

Picture	% S Jdgmts	% F Jdgmts	Picture	% S Jdgmts	% F Jdgmts	Picture	% S Jdgmts	% F Jdgmts	Picture	% S Jdgmts	% F Jdgmts
S1–O..	71	29	F21–O..	35	65	S1–Y..	46	54	F21–Y..	67	33
S2–O..	81	19	F22–O..	67	33	S2–Y..	45	55	F22–Y..	62	38
S3–O..	38	62	F23–O..	77	23	S3–Y..	61	39	F23–Y..	44	56
S4–O..	68	32	F24–O..	42	58	S4–Y..	32	68	F24–Y..	25	75
S5–O..	47	53	F25–O..	71	29	S5–Y..	55	45	F25–Y..	83	17
S6–O..	67	33	F26–O..	36	64	S6–Y..	58	42	F26–Y..	74	26
S7–O..	45	55	F27–O..	26	74	S7–Y..	41	59	F27–Y..	20	80
S8–O..	24	76	F28–O..	51	49	S8–Y..	29	71	F28–Y..	47	53
S9–O..	60	40	F29–O..	25	75	S9–Y..	46	54	F29–Y..	35	65
S10–O..	46	54	F30–O..	20	80	S10–Y..	48	52	F30–Y..	56	44
S11–O..	52	48	F31–O..	40	60	S11–Y..	82	18	F31–Y..	33	67
S12–O..	60	40	F32–O..	40	60	S12–Y..	32	68	F32–Y..	71	29
S13–O..	49	51	F33–O..	60	40	S13–Y..	49	51	F33–Y..	57	43
S14–O..	41	59	F34–O..	60	40	S14–Y..	65	35	F34–Y..	30	70
S15–O..	68	32	F35–O..	55	45	S15–Y..	45	55	F35–Y..	75	25
S16–O..	16	84	F36–O..	46	54	S16–Y..	45	55	F36–Y..	50	50
S17–O..	65	35	F37–O..	83	17	S17–Y..	31	69	F37–Y..	65	35
S18–O..	30	70	F38–O..	44	56	S18–Y..	49	51	F38–Y..	41	59
S19–O..	18	82	F39–O..	46	54	S19–Y..	34	66	F39–Y..	43	57
S20–O..	61	39	F40–O..	30	70	S20–Y..	36	64	F40–Y..	59	41
Av.....	50.4	49.6	Av.....	47.8	52.2	Av.....	46.6	53.4	Av.....	51.9	48.1

"Can untrained observers predict probable vocational success or failure on the basis of a photograph?"

The original 20 observers in this exercise showed a constant tendency to rate the men as successful, that is, they rated 7 men out of 10 successful whether the man was a

success or not. When the directions were changed in the check this tendency disappears and we find that there is clearly a chance distribution in the selection, 50 per cent. of each group (young men and older men) being judged as successful. We feel confident that any study conducted along similar lines will give the same result, namely, that vocational success or failure cannot be predicted or judged from photographs. We are also confident that it would be possible to select a group of photographs of successful men which would show a very high percentage in correctness of judgment or to select a group which would show a marked lack of accuracy. Much of the selection would be based on clothing, finish of the photograph, and the like, with but slight emphasis on the physiognomy itself.

The second question which this study proposed to attack was that of whether a group of untrained observers can judge from photographs whether or not an individual who has been following a particular vocation for 25 years has been successful in that vocation. As with the previous question the 'relatively successful' men were selected correctly 7 times out of 10, while the 'relatively unsuccessful' were called successful with about equal frequency. Several explanations might be offered. Our judges may have been rather uncritical and inclined to rate most well-dressed, elderly men as successful or they may have been influenced by the instructions. It may be that vocational failure marks one's physiognomy only when that failure is keenly felt by the individual; and most men refuse to admit failure. It may be that success in other things than one's vocation will overcome whatever effect the relative vocational unsuccess might have had in marking the face. Or it may be that only abject 'down-and-out' failure will leave its traces on the face. Whatever it is, we feel that the present study indicates that neither success or failure marks the static physiognomy of a photograph.

The evidence concerning the question of possible vocational selection and prediction from photographs is clean-cut. This group of untrained judges could not tell what

profession the young men would follow nor could they tell what profession the older men had been following for 25 years.

The believer in physiognomy might object that our judges were untrained whereas the trained judge could make accurate predictions and selections. He might further object that the physiognomy of the individuals indicated that they should have followed other professions for which their *type* was better fitted—that the 'relatively unsuccessful' men were unsuccessful because they had not followed the correct profession. To the first objection we would reply that no physiognomical treatise has yet given objective standards on which all physiognomists are agreed and which can be experimentally verified. Until these are given, to be a 'trained' physiognomist means little or nothing. To the second objection we would reply that our results indicate that men who had been successful are placed in as wide a variety of possible professions as those who are relatively unsuccessful. If it were true that physiognomical *type* determined success or failure, depending on whether the type matched the job, then we should have found the successful men grouped by the judges in only a few possible professions. This we did not find.

Physiognomists usually claim that a judgment of character or of traits of personality is based on certain specific features or facial relationships, *e.g.*, the length of the nose or the shape of the mouth and chin. We asked our judges to state the basis on which they made their judgment, giving them a list of possible bases and instructing them to use these reasons where they covered the case and when the list was not adequate to enter on the questionary, in the column headed 'remarks,' the real basis. The results which are summarized in Table IV have been noted above. Attention should be called to the frequency which 'hunch,' 'general impression' and 'facial expression' occur; namely, a total of 48.3 per cent. and 54.8 per cent. for the Older Men and for College Seniors, respectively. When we group the percentages for shape of face, appearance of eyes, length of nose, width of forehead, height of forehead, shape and size of ears, condition, neatness, etc., and position of head, which are the bases of specific judgments, we get 21.9 per cent. and 23.5 per cent. for the younger and older men, respectively. That is to say, approximately 1 judgment out of 4 is based upon specific traits, 2 out of 4 on 'hunch' and 1 out of 4 on a comparison or similarity of appearance to some known individual. It is evident that judgments of character and personality depend only to a minor degree on the perception of particular and specific features.

The last question raised concerned the certainty of judgment. We found our judges slightly more certain of their judgments of older men than of younger (34.8 per cent. as compared to 29.3 per cent.). Practically no difference in certainty existed between the average judgments of men who were successful and those who were unsuccessful. The range of certainty from one individual photograph to another was small. Thus there was no tendency for all judges to feel uncertain or certain about any one photograph. The correlation (rank-difference method) between the average certainty of judgment for the most successful group when young and

when old is − .12. The correlation between the per cent. of 'S' judgments for the same groups is − .11.

In general we may say that the present study lends absolutely no evidence toward the confirmation of the claims of the physiognomists. It offers no evidence which would indicate that the inclusion of a photograph in a letter of application or on a personnel sheet is of any positive value. This is an interesting point to the psychologist. All of us have a more or less firmly fixed conviction that we can judge a man's ability, personality, etc., from seeing his photograph. Whether this conviction is based on fact or based on something of the same sort of logic with which we justify our conviction that we can tell when some one is staring at us is conjectural. The writers are inclined to accept the latter explanation.

In an experiment of this type the accuracy of the judgments would no doubt have been increased to some extent if the observers had had a chance to see the subjects in person. A brief interview with each of the subjects would have considerably increased the accuracy of the decisions, for then those who were making the judgments would have had an opportunity to talk with the men. From the subject's conversation and manner of speaking the observers would doubtless have received some indication as to the degree of success he had achieved. This method should be very helpful, especially in distinguishing between extremes. Such traits as forcefulness, vigor, methods of speech, choice of diction, and indications of range of knowledge, which are not revealed in the pictures, would have aided in distinguishing between those men who are preëminently successful and those who are relatively unsuccessful.

Summary

Photographs of a group of men who had been relatively successful and a group who had been relatively unsuccessful in their profession were presented to untrained judges who were asked to state on the basis of the photograph whether the man would be or had been successful; what vocation he should or had followed; together with the basis of the judgment made and its certainty. The following results were obtained.

1. This group of untrained judges failed to make diagnostic judgments of either vocational success or of its lack. (There was a tendency to rate a man as successful rather than unsuccessful, especially when no definite directions to the contrary had been given.)

2. Since in predicting and judging success the observers placed far too many subjects among the 'most successful,' it seems that the judges cannot really discriminate well between the two qualities and are inclined to over-emphasize the number of successes.

3. It is practically impossible to determine what vocation a college senior should follow merely by examining his photograph. It is equally difficult to judge what vocation a man is pursuing who graduated from college approximately twenty-five years before the picture was taken.

4. In an experiment of this type the percentage of certainty that the observers feel concerning their judgments of success and failure is relatively low. The average percentage of certainty for a judgment in this study was 32 per cent.

5. Forty-eight per cent. of the reasons given in justification of the judgments may be classified as a 'general impression,' a 'hunch,' or the like; thirty-two per cent. as similarity to some acquaintance or specific person and twenty-two per cent. as definitely referring to some definite physiognomical feature or features.

References

1. Anderson, L. D. Estimating intelligence by means of printed photographs, *J. Appl. Psychol.*, 1921, 5, 152–155.
2. Blackford, K. M. H., & Newcomb, A. *Analyzing Character*, Alden, New York, 1916.
3. Gaskill, Fenton & Porter. Judging the intelligence of boys from their photographs, *J. Appl. Psychol.*, 1927, 11, 394–404.
4. Pintner, R. Intelligence as estimated from photographs, *Psychol. Rev.*, 1918, 25, 286–296.

(Manuscript received Dec. 23, 1927)

AN ELECTRO–KYMOGRAPH USED IN RECORDING VASO–MOTOR CHANGES

BY STANLEY B. LINDLEY

University of Kansas

The following pages describe an apparatus by which simultaneous records of heart action, breathing, maze circuits and time can be traced upon a continuous strip of moving paper. An investigation into the relationships between vaso-motor rhythms and rhythms of the voluntary neuro-muscular system, in progress at the University of Kansas, necessitated a means of obtaining continuous records up to an hour or two hours in length. A form of the electro-kymograph was adopted (see Fig. 1) with the use of rolls of ordinary wrapping paper in place of the usual smoked glazed paper. Another advantage offered by the electro-kymograph was that the recording pointers did not touch the paper, thus eliminating friction and critical adjustments of the tambour pointers. Heart beats of 2–3 cm in amplitude were readily obtained.

The working principle of the electro-kymograph lay in the passage of high tension currents between the pointers and drum. These currents penetrated the intervening strip of moving paper and produced dotted lines burned by the alternating sparks. A section of a typical record is shown in Figure 2.

The hand maze used in the experiment had no blind alleys but did have the sides of the path notched in such a manner as to force the subject to remain in the middle of the path. At the starting point of the maze was a lever which was struck by the stylus whenever the subject completed a full circuit of the path and returned to the starting point. Striking the lever opened a switch and broke a circuit on one of the recording pointers on the kymograph thus interrupting a straight line of holes on the record paper and providing a means by which the length of time in seconds per trial could be determined. The unoccupied arm of the subject was placed in a Lehmann plethysmograph which was suspended from the ceiling. A Marey pneumograph also was used. The pneumograph tambour and the tambour for the plethysmograph were mounted on a horizontal rod of glass which could be adjusted by a universal standard. Thus the recording pointers were easily kept at the proper distance from the moving strip of paper.

The electro-kymograph consisted of two units, the spark coils and the kymograph or recording instrument. A set of four Ford coils actuated by a current transformed and reduced from an ordinary alternating wall circuit produced the high tension current by which the recording pointers left the trail of their movements on the passing paper. The noise from the transformer and coils seemed not to produce any noticeable degree of distraction in the subjects, at least after the first few minutes of practice. A sound-proof booth for these components of the apparatus might, however, be advisable for certain kinds of work.

The wires from the generating unit were strung overhead on a frame-work in a manner designed to eliminate contact with other wires and to prevent contact with the experimenter. The four wires dropped vertically from the frame-work to the four pointers suspended over the kymographic drum. The ground wire from the coils

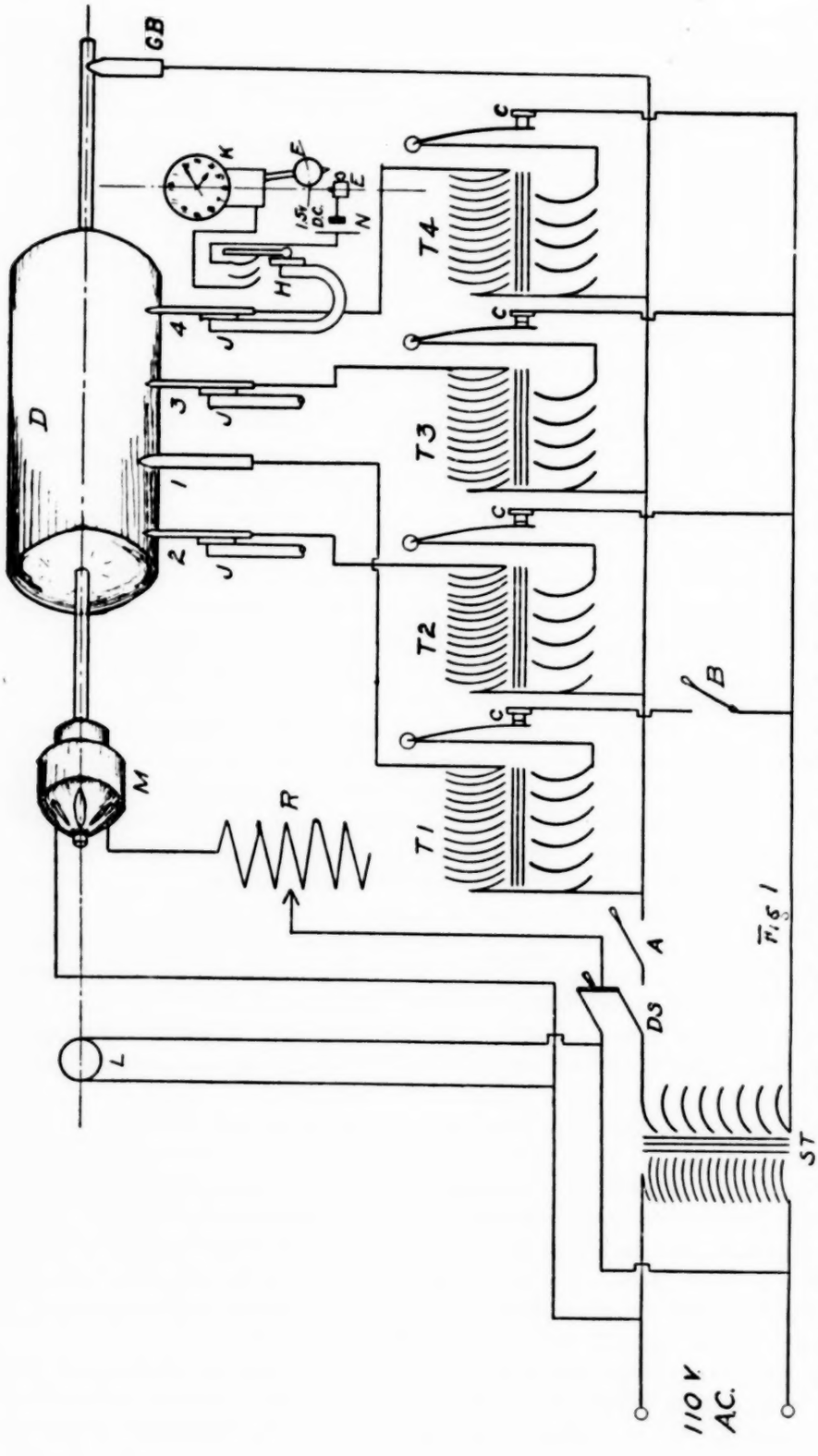

FIG. 1. *ST*, step-down transformer; *DS*, double-throw switch; *A*, single-throw switch; *B*, maze stop-switch; *C*, coil vibrator interrupter contacts; T_1, T_2, T_3, T_4, Ford coils; *L*, room lamp; *M*, motor; *R*, rheostat; 1, 2, 3, 4, kymograph pointers; *J*, *J*, *J*, tambours; *H*, magnetic timing tambour mechanism; *E*, mercury bulb; *F*, pendulum; *K*, clock; *GB*, ground brush; *D*, kymograph drum.

led to a brush by which the current reached the brass kymographic drum. The horizontal drum was turned by a small laboratory motor through a simple reduction and worm gear. The speed of the motor was regulated by a rheostat and fly-wheel.

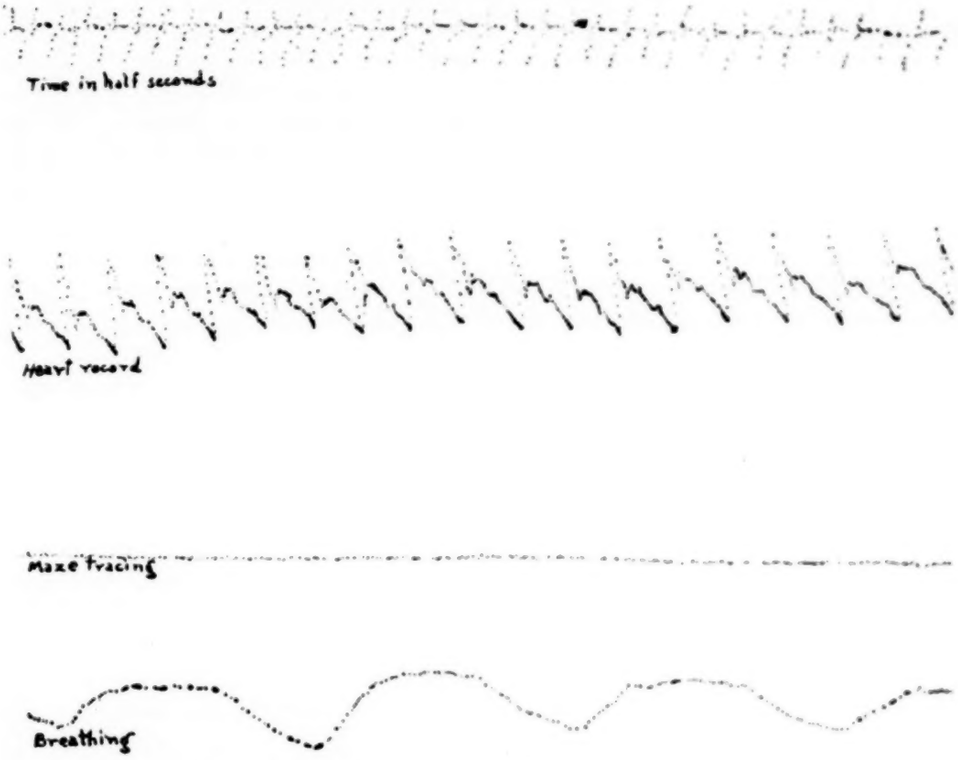

FIG. 2. Tracings made on cheap wrapping paper

The rapidly alternating current from the coils produces an intense spark between the pointers and the brass drum when the gap is of the proper width. The adjustment of the pointers was not difficult or critical and generally the gap was just wide enough to prevent the pointers touching the paper. There is a slight, but negligible, attraction between the drum and the pointers by reason of the passage of the current. The record produced by a very rapidly fluctuating pointer, in some instances, had to be traced over with pencil in order to make the line clearly visible; but barring an occasional bit of tracing there is no preparation of the recording paper either before or after the experiment.

A decrease in the efficiency of the coils was noticeable about every ten or fifteen minutes despite slight increases of voltage at the input. This is probably due to internal heating of the coils and may be remedied, as by a potentiometer in the primary circuit which could be adjusted in accordance with readings from a high frequency voltmeter in the secondary circuit.

To prevent short-circuiting the timing record was produced by means of a tambour actuated by a device to which it was connected by means of a rubber tube. A pendulum clock closed a mercury contact every half second and allowed a low voltage current momentarily to activate a small bell magnet which pressed the end of a lever upon the rubber head of a tambour. The impulse from the tambour through the rubber tube forced out the pointer of the timing tambour suspended over the revolving

recording drum. Thus every half-second a jog was produced in the timing line on the moving paper.

When used in conjunction with the problem for which it was constructed the kymograph recorded the breathing, pulse-rate, amplitude of the pulse, individual characteristics of the heart beat, changes in blood volume in the arm, and the number of circuits of the maze accomplished by the subject. The timing line provided a means by which each circuit of the maze could be recorded in half-second units. Through the same device the heart and breathing records could be timed.

A device by which the upper part of the subject's arm could be held immovable was required due to the necessity of preventing all movement of the plethysmograph. This bit of apparatus consisted of a wooden frame-work, securely fastened to the table, in which the upper arm was held tightly by adjustable straps. A rather complicated system of rods and clamps interposed between the wooden frame and the table permitted the adjustment of this arm-rest to suit the individual requirements of the subjects.

This electro-kymograph was built, for the greater part, from standard laboratory material and scraps from the work-shop. The spark coils were the only parts bought with the view of being used in the kymograph. The whole set of apparatus used in the experiment requires but little time and effort in the upkeep.

The author has in mind the possibility of constructing a very compact portable electro-kymograph which could find many practical uses.[1]

---

[1] The writer acknowledges the great aid which Professor Raymond H. Wheeler rendered him, both in presenting helpful suggestions and in providing material employed in the construction of the apparatus. He also wishes to thank Mr. Ian MacLaren for clearing up many of the technical problems involved in designing and developing the electrical features of the kymograph.

(Manuscript received Feb. 29, 1928)